PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:	WO 95/01098
A01N 63/00	A3	(43) International Publication Date:	12 January 1995 (12.01.95)

(21) International Application Number: PCT/US94/07252

(22) International Filing Date: 24 June 1994 (24.06.94)

(30) Priority Data:
08/083,948 28 June 1993 (28.06.93) US

(71) Applicant: MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventors: CORBIN, David, Richard; 14453 Brittania Drive, Chesterfield, MO 63017 (US). GREENPLATE, John, Thomas; 747 Sherwick Terrace, Manchester, MO 63021 (US). JENNINGS, Michael, Girard; Apartment 307, 14550 Rialto Drive, Chesterfield, MO 63017 (US). PURCELL, John, Patrick; 615 Charbray Drive, Ballwin, MO 63011 (US). SAMMONS, Robert, Douglas; P.O. Box 69, New Melle, MO 63365 (US).

(74) Agent: BOLDING, James, Clifton; Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CN, CZ, FI, GE, HU, JP, KE, KG, KR, KZ, LK, LV, MD, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
  With international search report.
  Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
16 February 1995 (16.02.95)

(54) Title: METHOD OF CONTROLLING INSECTS

(57) Abstract

3-Hydroxysteroid oxidase controls insects. Genes encoding for this enzyme may be cloned into vectors for transformation of plant-colonizing microorganisms or plants, thereby providing a method of controlling insect infestation.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5   A01N63/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5   A01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol.196, no.3, 15 November 1993, NEW YORK pages 1406 - 1403 J.P. PURCELL ET AL. 'Cholesterol oxidase: a potent insecticidal protein active against boll weevil larvae.' see the whole article	1-9

☐ Further documents are listed in the continuation of box C.   ☐ Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 December 1994	21.12.94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	DECORTE, D

Form PCT/ISA/210 (second sheet) (July 1992)

PCT

(51) International Patent Classification 6 : A01N 63/00	A2	(11) International Publication Number: WO 95/01098
		(43) International Publication Date: 12 January 1995 (12.01.95)

(21) International Application Number: PCT/US94/07252

(22) International Filing Date: 24 June 1994 (24.06.94)

(30) Priority Data:
08/083,948   28 June 1993 (28.06.93)   US

(71) Applicant: MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventors: CORBIN, David, Richard; 14453 Brittania Drive, Chesterfield, MO 63017 (US). GREENPLATE, John, Thomas; 747 Sherwick Terrace, Manchester, MO 63021 (US). JENNINGS, Michael, Girard; Apartment 307, 14550 Rialto Drive, Chesterfield, MO 63017 (US). PURCELL, John, Patrick; 615 Charbray Drive, Ballwin, MO 63011 (US). SAMMONS, Robert, Douglas; P.O. Box 69, New Melle, MO 63365 (US).

(74) Agent: BOLDING, James, Clifton; Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CN, CZ, FI, GE, HU, JP, KE, KG, KR, KZ, LK, LV, MD, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD OF CONTROLLING INSECTS

(57) Abstract

3-Hydroxysteroid oxidase controls insects. Genes encoding for this enzyme may be cloned into vectors for transformation of plant-colonizing microorganisms or plants, thereby providing a method of controlling insect infestation.

METHOD OF CONTROLLING INSECTS

FIELD OF THE INVENTION

This invention relates to a method of controlling lepidopteran insects by use of an enzyme which may be applied directly to the plant
5   or produced thereon by microorganisms or by genetically modifying the plant to produce the enzyme, and to genes, microorganisms, and plants useful in that method.

BACKGROUND OF THE INVENTION

10   The use of natural products, including proteins, is a well known method of controlling many insect pests. For example, endotoxins of *Bacillus thuringiensis (B.t.)* are used to control both lepidopteran and coleopteran insect pests. Genes producing these endotoxins have been introduced into and expressed by various plants, including cotton,
15   tobacco, and tomato. There are, however, several economically important insect pests that are not susceptible to *B.t.* endotoxins. There is also a need for additional proteins which control insects for which *B.t.* provides control in order to manage any development of resistance in the population.
20   It is therefore an object of the present invention to provide proteins capable of controlling lepidopteran insects and genes useful in producing such proteins. It is a further object of the present invention to provide genetic constructs for and methods of inserting such genetic material into microorganisms and plant cells. It is another object of the
25   present invention to provide transformed microorganisms and plants containing such genetic material.

SUMMARY OF THE INVENTION

It has been discovered that proteins that catalyze the oxidation of
30   3-hydroxysteroids, for example, cholesterol, will control lepidopteran insects. They cause mortality and stunting of larvae of lepidopteran insects. The enzymes may be applied directly to plants or introduced in other ways such as through the application of plant-colonizing micro-organisms or by the plants themselves, which have been transformed to
35   produce the enzymes.

2

3-Hydroxysteroid oxidases (E.C.1.1.3.6) are naturally produced by microorganisms such as *Streptomyces* sp., *Pseudomonas* sp., *Mycobacterium* sp., *Schizophyllum commune*, *Nocardia* sp., and *Rhodococcus* sp. [Smith et al., 1976, and Long et al., 1990.].

5   Preparations of enzymes from several different sources are available from Sigma Chemical Company, St. Louis, Missouri.

New *Streptomyces* genes that control the expression of 3-hydroxysteroid oxidase have been isolated and sequenced. These new genes or genes from other known producers of 3-hydroxysteroid oxidase

10   may be inserted into a transformation vector cassette which is used to transform plant-colonizing microorganisms which when applied to plants express the genes producing a 3-hydroxysteroid oxidase, thereby providing control of insects. Alternatively, genes which function in plants and encode the subject enzymes may be inserted into

15   transformation vector cassettes which may be incorporated into the genome of the plant, which then protects itself from attack by expressing the gene and producing a 3-hydroxysteroid oxidase. Additionally, the plant may also be transformed to co-express *B.t.* genes which express proteins for the control of other insects. Examples of

20   plants transformed to express *B.t.* genes are disclosed in European Patent Publication No. 0 385 962 (Fischhoff et al.).

In accomplishing the foregoing, there is provided, in accordance with one aspect of the present invention, a method of controlling insect infestation of plants comprising providing a 3-hydroxysteroid oxidase for

25   ingestion by the insect.

In accordance with another aspect of the present invention, there is provided a method of producing genetically transformed plants which express an amount of a 3-hydroxysteroid oxidase effective to control lepidioteran insects, comprising the steps of:

30   a)   inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising

(i)   a promoter which functions in plant cells to cause the production of an RNA sequence;

(ii)   a structural coding sequence that encodes for

35   3-hydroxysteroid oxidase; and

(iii)   a 3' non-translated region which functions in said plant
        cells to cause the addition of polyadenylate nucleotides to
        the 3' end of the RNA sequence,

    wherein said promoter is heterologous with respect to the
    structural coding sequence and wherein said promoter is
    operatively linked with said structural coding sequence, which is
    in turn operably linked with said non-translated region;

    b)   obtaining transformed plant cells; and

    c)   regenerating from the transformed plant cells genetically
        transformed plants which express an insecticidally effective
        amount of sterol oxidase.

        There is also provided, in accordance with another aspect of the
present invention, bacterial and transformed plant cells that contain
DNA comprised of the above-mentioned elements (i), (ii), and (iii).

        As used herein, the term "controlling insect infestation" means
reducing the number of insects which cause reduced yield, either through
mortality, retardation of larval development (stunting), or reduced repro-
ductive efficiency.

        As used herein, the term "structural coding sequence" means a
DNA sequence which encodes for a polypeptide, which may be made by
a cell following transcription of the DNA to mRNA, followed by
translation to the desired polypeptide.

DETAILED DESCRIPTION OF THE INVENTION

        3-Hydroxysteroid oxidases catalyze the oxidation of the
3–hydroxy group of 3-hydroxysteroids to produce ketosteroids and
hydrogen peroxide. They are capable of catalyzing the oxidation of
various 3-hydroxysteroids, such as, for example, cholesterol. Most of the
previously known 3-hydroxysteroid oxidases are called "cholesterol
oxidases" (enzymatically catalogued as E.C. #1.1.3.6) but cholesterol is
only one of the 3-hydroxysteroid substrates. The use of all 3-hydroxy-
steroid oxidases and the genes encoding such proteins, for the purpose of
controlling insects, are within the scope of the present invention.

        3-Hydroxysteroid oxidases are commercially available for use as
reagents for serum cholesterol assays. For example, Sigma Chemical

Company, St. Louis, Missouri, offers three such 3-hydroxysteroid oxidases (denominated as cholesterol oxidases), one from a *Streptomyces* sp., one from a *Pseudomonas fluorescens*, and one from a *Brevibacterium*. Two other sources of 3-hydroxysteroid oxidase, two streptomycetes

5   denominated A19241 and A19249, each of which produce a 3-hydroxy-steroid oxidase, have been isolated. The organisms were collected in Madagascar. When these organisms were cultured according to usual methods the culture filtrates were found to affect insect larvae as described below.

10   A seed culture of A19249 was started in 55 mL sterile Tryptone-Yeast Extract broth, pH 6.8, in a 250 mL shake flask. The seed was agitated at 250 rpm on a rotary shaker for 3 days at 30 °C. A New Brunswick Bioflo II Bioreactor with a 2 L working volume was filled with "medium 202" ($MgSO_4 \cdot 7H_2O$ 2 g/L, KH_2PO_4 0.5 g/L, NaCl 0.5 g/L,

15     $CaCO_3$ 1 g/L, $ZnSO_4 \cdot H_2O$ (1 mg/mL stock) 5 mL/L, 100 mM FeEDTA 0.5 mL/L, Soluble Starch 5 g/L, Dextrose 2.5 g/L, Malt Extract 2.5 g/L, Soytone 5 g/L). The pH was adjusted to 6.5 with 2.5 N NaOH or 1 N HCl, and 1 mL/L of P2000, an antifoam agent was added. The bioreactor was sealed and autoclaved for 25 min at 250 °C. The seed, at

20   3 days growth, was used to inoculate the fermentor at 2% or 40 mL. The fermentation took place at 30 °C with an airflow of 1 L/min and agitation running at 500 rpm. The fermentation was harvested after 40 h.

           Each of these enzymes has demonstrated control of insects as

25   shown below. The *P. fluorescens* 3-hydroxysteroid oxidase is immunologi-cally distinct from the *Streptomyces* enzymes, but it also controls insects.

           Other organisms producing 3-hydroxysteroid oxidases of the present invention may be identified by assaying culture filtrates or

30   individual proteins for 3-hydroxysteroid oxidase activity using a spectro-photometric assay, described below, which measures hydrogen peroxide production in the presence of a 3-hydroxysteroid, for example, cholesterol [Gallo, 1981].

BIOEFFICACY ASSAYS

<u>Larval Bioassay</u>

Lepidopteran larvae were tested on artificial agar-based diet
5   (Marrone et al., 1985) treated with the indicated amount of the A19249
3-hydroxysteroid oxidase (cholesterol oxidase) for six days. The results
are shown in Table 1. Percent stunting is defined as

<u>larval weights (Hepes control) - larval weights (cholesterol oxidase)</u>
x100%
10   larval weights (Hepes control)

An extended test was performed with tobacco budworm larvae to
test the effect of the stunting noted in the six-day test. Tobacco
budworm eggs were added to artificial diet (as described above)
containing either buffer or 100 ppm A19249 3-hydroxysteroid oxidase
15   (cholesterol oxidase). After seven days, some mortality as compared to
the controls was noted. Surviving larvae were moved to fresh diet
(control or treated, as appropriate) every seven days. Percent mortality.
(corrected for control mortality) is reported for the 7, 10, and 14 day
periods in Table 2. The corrected number of larvae was 23. By the 27th
20   day only two larvae were alive; both were very stunted and never
pupated. 92% of the control larvae survived and all survivors pupated
by day 27. This experiment demonstrates that 3-hydroxysteroid oxidase
essentially arrests the development of tobacco budworm larvae.

25   <u>Table 1</u>

<u>Insect</u>	<u>Stage</u>	<u>Dose-µg/mL</u>	<u>Stunting</u>
tobacco budworm	egg/lv	30	0
	lv	100	86%
corn earworm	lv	50	0
	lv	100	35%
fall army worm	lv	30	0
tobacco hornworm	lv	30	0
	lv	100	30%
pink bollworm	lv	50	0
	lv	100	30%

6

European cornborer	lv	50	0
	lv	100	46%
beet armyworm	lv	100	76%
black cutworm	lv	100	68%

Table 2

Interval (days)	Percent Mortality
7	20
10	61
14	80

MODE OF ACTION STUDIES

The following studies show that 3-hydroxysteroid oxidase has a direct effect on the insect itself and that the activity demonstrated in the experiments described above cannot be attributed to the enzymes effect on the insect's diet, for example by sterol depletion. Boll weevil and lepidopteran larvae are most susceptible to the enzyme. It is believed that this specificity is due to the effect of 3-hydroxysteroid oxidase on the midgut of the insect as explained in more detail below. Other insects with similar midgut physiologies may also be controlled by 3-hydroxysteroid oxidase. In addition, 3-hydroxysteroid oxidases other than those tested and reported herein may control a different spectrum of insects with different midgut physiologies.

Cotton Seed Diet Assay

Treatment diet was made by mixing 30 g of Pharmamedia™ (Traders Protein) cottonseed flour into 170 mL of a 1.6% agar solution at 50 °C, containing 0.13% propionic acid, 0.014% phosphoric acid, and 30 mg each of streptomycin sulfate and chlor-tetracycline. Before mixing, 10% KOH was used to adjust the pH to 6.2. Pharmamedia™ is a flour made up of cottonseed embryos. The diet was incubated in a water bath at 40 °C. Dilutions of the Sigma *Streptomyces* 3-hydroxysteroid oxidase (cholesterol oxidase) were incorporated into the diets. Tobacco budworm larvae are 68% stunted when exposed to 3-hydroxysteroid oxidase (100 ppm) in cottonseed diet.

7

<u>Cotton Callus Diet Assay</u>

Cotton callus assays were conducted using Coker 312 cotton callus and a 96 well insect diet tray. Each well contained 0.5 ml of gelled

5   2% agar (containing 0.13% propionic acid and 0.014% phosphoric acid) that was covered with a filter paper disc. A piece of callus (approximately 150 mg) was placed in each well. For each replicate, 25 µl of a stock solution of 3-hydroxysteroid oxidase (1.25 mg/ml stock) or 25 µl of 25 mM Hepes buffer (pH 7.5) was pipetted onto each piece of

10   callus. Tobacco budworm eggs (4-8 eggs/well, 0-12 hr from hatching) in 0.15% agar were pipetted onto the filter disc next to the callus. Individual larvae were transferred to freshly treated callus samples every three days. Larvae were weighed after 10 days.

Stunting of larvae was observed in the treated wells, although

15   reduced as compared to the artificial diet or the cottonseed diet. This reduced level of stunting with cotton callus is probably due to the pipetted enzyme solution not completely covering the callus tissue and the loss of some enzyme onto the filter paper.

20   <u>Pre-ingestion Effect on Diet</u>

There are no lethal or stunting effects from feeding larvae a diet sample that was incubated for one week with a 3-hydroxysteroid oxidase and then heated to 80 °C to denature the enzyme prior to using it in the above assay. This further demonstrates that the mode of action of

25   3-hydroxysteroid oxidase is not dependent on sterol depletion of the nutrient source but that the enzyme is directly active upon the insect.

<u>Microscopy Studies</u>

Tobacco budworm larvae were reared from eggs on artificial diet

30   containing 100 µg/ml of 3-hydroxysteroid oxidase or Hepes buffer control. After ten days, midgut epithelia were dissected and processed for microscopy as described by Purcell et al., 1993. The midguts from treated larvae exhibited a variety of histological and ultrastructural alterations as compared with Hepes buffer-treated controls. 3-Hydroxy-

35   steroid oxidase induced a lateral constriction and apical-basal elongation

8

of the cells in the midgut epithelium, resulting in a widening of the inter-
cellular spaces and the formation of an irregular and convoluted lumenal
surface. The cellular attenuation was also characterized by localized
apical cytoplasmic blebbing and microvillar denudation. This denudation
5   occasionally generated detached cytoplasmic fragments in the lumen.
Localized cytolysis was observed, but the epithelium as a whole
remained intact. These observations suggest that exposure to 3-
hydroxysteroid oxidase compromises the integrity of the apical plasma
membrane which results in a secondary generalized cellular
10   hypersensitivity to osmotic swelling and lysis.

Cholesterol is required for the integrity and normal function of
virtually all cellular membranes. A reduction in the availability of
cholesterol for incorporation into and maintenance of cellular
membranes due to metabolism by 3-hydroxysteroid oxidase could there-
15   fore be one possible mechanism for toxicity of this agent. However, the
observations described above show the most dramatic structural
alterations are localized to the tips of the midgut cells in contact with
the 3-hydroxysteroid oxidase in the gut lumen. This suggests that 3-
hydroxysteroid oxidase may act directly to alter cholesterol already
20   incorporated into the membrane. In addition, diet treatment studies
described above demonstrated that the diet was not altered prior to
ingestion in a manner which compromises tobacco budworm growth,
again suggesting a direct effect of the enzyme on the insect.

25   Mode of Action Theory

While not being bound by this theory, it is believed that the 3-
hydroxysteroid oxidase enzyme stunts the growth of lepidopteran larvae
by some action in the gut after ingestion. The bioassay and
morphological data show that there was clearly a pathological condition
30   resulting from ingestion of diet containing 3-hydroxysteroid oxidase.

ENZYME IDENTIFICATION

The active proteins from the Madagascar *Streptomyces* micro-
organisms were isolated, purified, partially sequenced, and identified as
35   3-hydroxysteroid oxidases.

Protein Isolation

Each culture filtrate was purified by first sizing on YM10 membranes (Amicon) to a [>10 kDa] fraction, followed by multiple chromatography runs on an FPLC Mono Q HR10/10 (Pharmacia LKB, Piscataway, NJ) column. For chromatography on the Mono Q column, the samples were loaded on the column in 25 mM Hepes pH 7.5 and eluted with a gradient to 1.0 M KCl in 25 mM Hepes pH 7.5. Fractions were collected and aliquots of each were filtered through 0.2 μ Acrodisc syringe tip filters. Each was tested in an insect assay. Aliquots of insecticidally active fractions were electrophoresed on SDS-PAGE [Laemmli, 1970] using a Daiichi Double Gel Device and 10-20% mini-gel. Proteins were visualized by silver staining using Daiichi silver stain reagent kit. The active enzymes of the present invention, isolated from the novel microorganisms, were found to be a 52.5 kDa protein.

Amino Acid Sequences

An SDS-PAGE gel of the protein produced by *Streptomyces* A19241, isolated as above, was blotted onto PVDF paper (Immobilon, Millipore) using the protocol of Matsudaira [Matsudaira, 1987]. The N-terminus was sequenced using automated Edman degradation chemistry. A gas phase sequencer (Applied Biosystems, Inc.) was used for the degradation using the standard sequencer cycle. The respective PTH-aa derivatives were identified by reverse phase HPLC analysis in an on-line fashion employing a PTH analyzer (Applied Biosystems, Inc.) fitted with a Brownlee 2.1 mm i.d. PTH-C18 column. For internal sequences, digestions were carried out on purified 3-hydroxysteroid oxidase from A19249 using trypsin (TPCK-treated, from Worthington Biochemicals Corp., Freehold, NJ). Fragments were then purified by reverse phase HPLC and sequenced in an N-terminal fashion.

The resulting partial sequences were compared to known proteins and a strong (71%) homology was found with the reported fourteen amino acid sequence at the N-terminus of a 3-hydroxysteroid oxidase isolated from a *Streptomyces* species [Ishizaki et al., 1989]. The reported enzyme has an M_r of 54.9 kDa which agrees well with the M_r of 52.5 kDa of the isolated enzyme.

Six internal fragments of the purified enzyme from A19249, also
having homology to six regions of the reported enzyme, were sequenced.
The fragments had 95, 76, 64, 58, 89, and 100 percent sequence
identities.

5

Amino Acid Composition Determination and Comparison

The amino acid composition of the 3-hydroxysteroid oxidase
produced by A19249 was determined and compared with the
composition of the reported *Streptomyces* enzyme. The samples were
10   subjected to acid hydrolysis (6N HCl, vapor phase hydrolysis using a
Water's Picotag workstation, 24 hr, 110 °C.). All analyses were
performed after post-column derivation of the hydrolysates using
ninhydrin [Moore et al., 1963]. A Beckman Model 6300 Auto analyzer
was employed for the actual determinations. The S delta n/N statistic is
15   used to compare two compositions in order to make a prediction about
their relatedness. The formula for the statistic is:

$$1/2 \, \Sigma \, (n_{Ai} - n_{Bi})^2/N$$

where A is one composition, B is the other composition, i is each amino
acid, n is the number of each amino acid, N is the total number of amino
20   acids in the protein. If S delta n/N is <0.42, then there is a greater than
95% chance that the proteins are related. The smaller the value, the
more closely the determined compositions match.

The S delta n/N statistic for the A19249 protein compared to the
reported enzyme is 0.36, indicating that the two are highly related.

25

3-Hydroxysteroid Oxidase Assay

The identity of the enzyme was confirmed by testing its ability to
oxidize a 3-hydroxysteroid, specifically cholesterol. The enzyme is added
to a reagent mixture comprising horseradish peroxidase (20 U/mL),
30   phenol (14 mM), 4-amino antipyrine (0.82 mM), Triton® X-100 (0.05%)
and phosphate buffer (100 mM, pH 7). The sterol in isopropanol is then
added and the absorbance at 500 nm monitored. One unit of activity is
defined as the amount of enzyme required to oxidize 1 µmole of sterol per
minute at 20 °C.
35   The activity levels of the enzymes are reported in Table 12 for 3-

hydroxysteroids representative of various classes of 3-hydroxysteroids.
The enzyme sources are as follows:

        1 = A19249

        2 = A19241

5   3 = Sigma *Streptomyces*

        4 = Sigma *Pseudomonas*

Table 3
Relative Rate for Enzymes

Sterol	1	2	3	4
10 cholesterol	100	100	100	100
dihydrocholesterol	56	56	59	69
dehydrocholesterol	13	12	7	47
lathosterol	28	34	27	71
stigmasterol	22	28	11	21
15 sitosterol	88	65	49	50
campesterol*	65	64	45	49
fucosterol	22	20	12	68
lanosterol	<1	<1	<1	1
ecdysone	<1	<1	<1	<1
20 20-OH ecdysone	<1	<1	<1	<1

*65/35 mixture of campesterol and dihydrobrassicasterol

Immunological Comparison of Enzymes

    The Sigma *Streptomyces* enzyme is immunologically related to the
25 3-hydroxysteroid oxidases produced by the isolates of the present
invention, numbers A19241 and 19249, as demonstrated by Western
blotting [Burnette et al., 1981] using polyclonal antisera generated
against the Sigma *Streptomyces* enzyme. The antisera recognized both
enzymes produced by the isolates. The 3-hydroxysteroid oxidase from *P.*
30 *fluorescens* was not recognized by the antisera. This demonstrates that
immunologically distinct 3-hydroxysteroid oxidases are toxic to insects.

GENETIC IDENTIFICATION

    The 3-hydroxysteroid oxidase gene was isolated from one of the
35 *Streptomyces* microorganisms isolated in Madagascar and its sequence

12

determined.

Cloning of the 3-Hydroxysteroid Oxidase Gene from A19249

As discussed above, peptide sequences of purified 3-hydroxysteroid

5   oxidase from A19249 were obtained for four regions of the protein. These
peptide sequences were compared to a database of known protein
sequences, and this comparison revealed that the A19249 protein showed
a high degree of homology to a known 3-hydroxysteroid oxidase from
Streptomyces [Ishizaki]. Comparing the A19249 peptide sequences to this

10   known protein sequence, these peptides were assigned to their likely
positions in the A19249 protein sequence. The sequence derived from the
intact 3-hydroxysteroid oxidase from A19249 corresponded to a region
near the N-terminus of the secreted form of the enzyme from the pub-
lished sequence. From this it was concluded that the A19249 N-terminal

15   peptide sequence was also likely to correspond to a "mature" secreted
form of the protein lacking its putative secretory signal sequence. This
was later confirmed by the DNA sequence analysis of the A19249 gene
(see below). Three peptides were used to construct hybridization probes
for isolation of the A19249 3-hydroxysteroid oxidase gene. Peptide N2

20   (SEQ ID NO:1) corresponded to N-terminal amino acids 29 - 43 of the
known mature protein sequence (sequence without the putative signal
peptide); peptide C1 (SEQ ID NO:2) to amino acids 434 - 449 of the
mature protein sequence; and peptide C2 (SEQ ID NO:3) to amino acids
464 - 475 of the mature protein sequence.

25   N2 (SEQ ID NO:1):
         ValSerThrLeuMetLeuGluMetGlyGlnLeuTrpAsnGlnPro
C1 (SEQ ID NO:2):
         AlaPheAlaAspAspPheCysTyrHisProLeuGlyGlyCysValLeu
C2 (SEQ ID NO:3):   AsnLeuTyrValThrAspGlySerLeuIleProGly

30   Based on these peptide sequences, three long nondegenerate
oligonucleotides, corresponding to 3-hydroxysteroid oxidase peptide
sequences from A19249, were designed using *Streptomyces* preferred
codons. The oligonucleotides N2 (SEQ ID NO:4), C1 (SEQ ID NO:5), and
C2 (SEQ ID NO:6) correspond to the peptides N2, C1, and C2 described

35   above.

N2 Probe (SEQ ID NO:4):

    gtgtccaccctgatgctggagatgggccagctgtggaaccagccc

C1 Probe (SEQ ID NO:5):

    gccttcgccgacgacttctgctaccacccgctcggcggctgcgtcctg

5 C2 Probe (SEQ ID NO:6):   aacctctacgtgaccgacggttcgctgatcccgggt

      Probes N2 (SEQ ID NO:4), C1 (SEQ ID NO:5), and C2 (SEQ ID NO:6) were all used as hybridization probes on Southern blots of A19249 genomic DNA. All three probes hybridized to the same 2.2 kb band in BamHI digested DNA, but N2 (SEQ ID NO:4) hybridized to a different

10 fragment than C1 (SEQ ID NO:5) and C2 (SEQ ID NO:6) did in SalI and BglII digests. This indicated that SalI and BglII cut within the coding sequence of the 3-hydroxysteroid oxidase gene from A19249, which was confirmed by DNA sequence analysis.

      The 3-hydroxysteroid oxidase gene from A19249 was isolated using

15 the three synthetic oligonucleotides as hybridization probes on a library of DNA fragments of A19249 DNA in a lambda phage vector. A library was made in lambda EMBL3 using partial-digest Mbo1 DNA fragments of A19249. These probes were used to screen approximately 72,000 lambda phage plaques from the primary library. Primary plaque screening was

20 performed using N2 (SEQ ID NO:4) plus C2 (SEQ ID NO:6). A total of 12 recombinant plaques that hybridized to the N and C-terminal probes were picked and purified by a second round of hybridization screening with probes N2 (SEQ ID NO:4) and C2 (SEQ ID NO:6). Southern blot analysis revealed that, in five of six lambda clones analyzed, a 2.2 kb BamHI frag-

25 ment hybridized to both the N and C-terminal probes. This result confirmed the earlier Southern hybridization analysis that indicated a 2.2 kb BamHI fragment contained the 3-hydroxysteroid oxidase gene. This 2.2 kb DNA fragment was cloned into plasmid vector pUC18 [Yanisch-Perron et al., 1985] in both orientations for further analysis. Restriction mapping

30 showed that there were internal SalI and BglII sites as predicted by the Southern hybridization analysis. These sites are also conserved compared to the published 3-hydroxysteroid oxidase gene sequence. The BamHI fragment was further subcloned into four fragments for direct DNA sequencing.

35

14

Sequence Analysis of the 3-Hydroxysteroid Oxidase Gene

A total of 1865 nucleotides of DNA sequence from the 2.2 kb
BamHI fragment were determined by direct DNA sequence analysis of
subclones of this fragment using the dideoxy chain termination method.

5   This sequence is identified as SEQ ID NO:7. This DNA sequence contains
noncoding flanking regions at both the 3' and 5' ends. Analysis of this
DNA sequence revealed a single long open reading frame that encodes a
secretory signal peptide and the mature 3-hydroxysteroid oxidase protein
of 43 and 504 amino acids, respectively. It is 84.37% identical to the

10 published 3-hydroxysteroid oxidase nucleotide sequence. The derived
amino acid sequence is 81.685% identical to the published 3-hydroxy-
steroid oxidase sequence. It is identified as SEQ ID NO: 8. Examination of
the A19249 DNA sequence and comparison to the N-terminal amino acid
sequence of intact 3-hydroxysteroid oxidase from A19249 revealed that

15 the A19249 gene encoded a protein that includes a signal peptide
sequence, which is apparently cleaved during secretion of the protein from
the cells. Thus the N-terminus of the mature protein from A19249 begins
with Ser-Gly-Gly-Thr-Phe, identified as SEQ ID NO:12.

20   GENETIC TRANSFORMATION

A 3-hydroxysteroid oxidase gene can be isolated from novel organ-
isms or may be obtained from known sources, such as the *Rhodococcus* sp.
described by Long et al., in WO 90 05,788. This gene may then be used to
transform bacterial cells or plant cells to enable the production of 3-

25 hydroxysteroid oxidase and carry out methods of this invention. Examples
of how this may be done with the gene of A19249 are given below.

Mutagenesis of the A19249 Gene

In order to incorporate the A19249 gene into vectors appropriate

30 for expression of the 3-hydroxysteroid oxidase in heterologous bacterial or
plant hosts, it was necessary to introduce appropriate restriction sites
near the ends of the gene. The goals of this mutagenesis were to create
cassettes that included the protein coding sequence with minimal non-
coding flanking sequences and to incorporate useful restriction sites to

35 mobilize these cassettes. Cassettes were designed that would allow mobili-

zation of the intact coding sequence including the signal peptide or just the mature coding sequence. To incorporate these cassettes into appropriate bacterial or plant expression vectors, an NcoI restriction site was engineered at the N-terminus of the intact protein sequence or at the

5   N-terminus of the mature protein sequence. A BamHI site was engineered just after the termination codon of the intact coding sequence. Three mutagenesis primers were designed to create these cassettes, as shown below. Mutagenesis with primer Chossn (SEQ ID NO:9) substituted three amino acids (MAT) for valine and asparagine at the N-

10   terminus of the signal peptide of the intact protein and Chomnr (SEQ ID NO:10) added two amino acids (MA) at the N-terminus of the mature protein. This was necessary to allow incorporation of the NcoI restriction site. Mutagenesis with primer Cho3br (SEQ ID NO:11) incorporated a BamHI site at the 3' end of the coding sequence. Primers Chomnr and

15   Cho3br were used to direct formation of the antisense strand of DNA.
      Chossn (SEQ ID NO:9): CTCAGGAGCACCATGGCGACCGCACAC
             (NcoI site underlined)
      Chomnr (SEQ ID NO:10): GTGCCGCCGGAGGCCATGGGGGCGGTGGC
             (NcoI site underlined)

20   Cho3br (SEQ ID NO:11):
      GCCCCGCCCGTCGGATCCGTCAGGAACCCG (BamHI site underlined)
      The resulting modified sequences were identified as SEQ ID NO:13
      encoding for the intact protein and SEQ ID NO:14 for the mature protein.

25   <u>Expression of 3-Hydroxysteroid Oxidase in *E. coli*</u>
             The NcoI-BamHI fragments containing either the intact protein coding sequence or the mature protein coding sequence were inserted into a vector designed for protein expression in *E. coli*, vector pKK233-2 (Pharmacia LKB, Piscataway, NJ). pKK233-2 contains the IPTG-

30   inducible trc promoter. The vector containing the intact (full length) protein coding sequence as modified (SEQ ID NO:13) is designated pMON20909. The vector containing the mature protein coding sequence as modified (SEQ ID NO:14) is designated pMON20907. *E. coli* XL1 Blue cells (Statagene, San Diego, CA) modified with pMON20909 expressed 3-

35   hydroxysteroid oxidase at higher levels of enzymatic activity than cells

modified with pMON20907. The protein was extracted and purified from 4
1 of IPTG-induced *E. coli* containing pMON20909. The soluble fraction
from sonicated bacterial lysate was concen-trated and dialyzed, and then
partially purified by Mono Q chromatography to yield 11 units of 3-hydroxy-
5 steroid oxidase activity. Western blot analysis indicates that the signal
sequence of the intact protein is cleaved in *E. coli*, but the exact site of
cleavage was not determined. Analysis of the recovered protein showed a
five-fold reduction in enzymatic activity relative to the A19249 protein,
but the loss has not been explained by DNA sequencing which found no
10 alterations that would explain loss of enzymatic activity in plant
protoplasts or *E. coli*.

The recovered protein was tested against tobacco budworm and
resulted in 88% stunting at a dose of 100 µg/ml.

15 <u>Expression of 3-Hydroxysteroid Oxidase in Plant Colonizing Bacteria</u>

To control insects, it may be desirable to express 3-hydroxysteroid
oxidase in plant colonizing bacteria, and then apply this bacteria to the
plant. As the insect feeds on the plant, it ingests a toxic dose of 3-hydroxy-
steroid oxidase produced by the plant colonizers. Plant colonizers can be
20 either those that inhabit the plant surface, such as *Pseudomonas* or *Agro-
bacterium* species, or endophytes that inhabit the plant vasculature such
as *Clavibacter* species. For surface colonizers, the 3-hydroxysteroid
oxidase gene may be inserted into a broad host range vector capable of
replicating in these Gram-negative hosts. Examples of these such vectors
25 are pKT231 of the IncQ incompatibility group [Bagdasarian et al., 1981]
or pVK100 of the IncP group [Knauf, 1982]. For endophytes the 3-hydroxy-
steroid oxidase gene can be inserted into the chromosome by homologous
recombination or by incorporation of the gene onto an appropriate trans-
poson capable of chromosomal insertion in these endophytic bacteria.
30

<u>Plant Gene Construction</u>

The expression of a plant gene which exists in double-stranded DNA
form involves transcription of messenger RNA (mRNA) from one strand
of the DNA by RNA polymerase enzyme, and the subsequent processing
35 of the mRNA primary transcript inside the nucleus. This processing

involves a 3' non-translated region which adds polyadenylate nucleotides to the 3' end of the RNA. Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the "promoter." The promoter region contains a sequence of bases that signals RNA polymerase to

5 associate with the DNA and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding strand of RNA.

A number of promoters which are active in plant cells have been described in the literature. Such promoters may be obtained from plants

10 or plant viruses and include, but are not limited to, the nopaline synthase (NOS) and octopine synthase (OCS) promoters (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the cauliflower mosaic virus (CaMV) 19S and 35S promoters, the light-inducible promoter from the small subunit of ribulose 1,5-bisphosphate carboxylase

15 (ssRUBISCO, a very abundant plant polypeptide), and the Figwort Mosaic Virus (FMV) 35S promoter. All of these promoters have been used to create various types of DNA constructs which have been expressed in plants (see e.g., PCT publication WO 84/02913).

The particular promoter selected should be capable of causing

20 sufficient expression of the enzyme coding sequence to result in the production of an effective amount of 3-hydroxysteroid oxidase. A preferred promoter is a constitutive promoter such as FMV35S. It has been observed to provide more uniform expression of heterologous genes in the flowering portions of plants. Use of such a promoter with the 3-hydroxysteroid

25 oxidase gene may provide greater protection of cotton bolls and squares from insect damage, than other promoters.

The promoters used in the DNA constructs (i.e. chimeric plant genes) of the present invention may be modified, if desired, to affect their control characteristics. For example, the CaMV35S promoter may be

30 ligated to the portion of the ssRUBISCO gene that represses the expression of ssRUBISCO in the absence of light, to create a promoter which is active in leaves but not in roots. The resulting chimeric promoter may be used as described herein. For purposes of this description, the phrase "CaMV35S" promoter thus includes variations of CaMV35S promoter,

35 e.g., promoters derived by means of ligation with operator regions, random

or controlled mutagenesis, etc. Furthermore, the promoters may be altered to contain multiple "enhancer sequences" to assist in elevating gene expression. Examples of such enhancer sequences have been reported by Kay et al. (1987).

5   The RNA produced by a DNA construct of the present invention also contains a 5' non-translated leader sequence. This sequence can be derived from the promoter selected to express the gene, and can be specifically modified so as to increase translation of the mRNA. The 5' non-translated regions can also be obtained from viral RNA's, from

10   suitable eukaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs wherein the non-translated region is derived from the 5' non-translated sequence that accompanies the promoter sequence. As shown below, a plant gene leader sequence which is useful in the present invention is the petunia heat shock protein

15   70 (Hsp70) leader. [Winter et al.]

As noted above, the 3' non-translated region of the chimeric plant genes of the present invention contains a polyadenylation·signal which functions in plants to cause the addition of adenylate nucleotides to the 3' end of the RNA. Examples of preferred 3' regions are (1) the 3' tran-

20   scribed, non-translated regions containing the polyadenylate signal of *Agrobacterium* tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene and (2) plant genes like the soybean 7s storage protein genes and the pea ssRUBISCO E9 gene. [Fischhoff et al.]

25   Plant Transformation and Expression

A chimeric plant gene containing a structural coding sequence of the present invention can be inserted into the genome of a plant by any suitable method. Suitable plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those

30   disclosed, e.g., by Herrera-Estrella (1983), Bevan (1983), Klee (1985) and EPO publication 0 120 516 (Schilperoort et al.). In addition to plant transformation vectors derived from the Ti or root-inducing (Ri) plasmids of *Agrobacterium*, alternative methods can be used to insert the DNA constructs of this invention into plant cells. Such methods may involve,

35   for example, the use of liposomes, electroporation, chemicals that increase

free DNA uptake, free DNA delivery via microprojectile bombardment, and transformation using viruses or pollen.

A particularly useful Ti plasmid cassette vector for transformation of dicotyledonous plants is pMON11782. The expression cassette
5 pMON11782 consists of the FMV35S promoter, the petunia Hsp70 5' untranslated leader, and the 3' end including polyadenylation signals from the pea ssRUBISCO E9 gene. Between the leader and the 3' polyadenylation signals is a multilinker containing multiple restriction sites, including a BamHI site for the insertion of genes. pMON11782 also
10 contains a HindIII site before the promoter sequence.

The remainder of pMON11782 contains a segment of pBR322 (New England Biolabs, Beverly, MA) which provides an origin of replication in *E. coli*; the oriV region from the broad host range plasmid RK1 which allows replication in *Agrobacterium* strain ABI; the streptomycin-
15 spectinomycin resistance gene from Tn7; and a chimeric NPTII gene, containing the CaMV35S promoter and the nopaline synthase (NOS) 3' end, which provides kanamycin resistance in transformed plant cells.

Another particularly useful Ti plasmid cassette vector is pMON17227. This vector is described by Barry et al. in WO 92/04449
20 and contains a gene encoding an enzyme conferring glyphosate resistance which is an excellent selection marker gene for many plants.

Transient Expression of 3-Hydroxysteroid Oxidase in Tobacco Plants

Both 3-hydroxysteroid oxidase gene cassettes, that is the gene
25 encoding intact protein with the signal sequence and that encoding only the mature protein, each modified at the N-terminus as described above, were mobilized as NcoI-BamHI fragments and inserted into a transient expression vector that had been cut with NcoI and BamHI. A transient expression vector is a simple plasmid containing a plant promoter with a
30 5' nontranslated leader, a 3' nontranslated polyadenylation sequence, and between them a multilinker having multiple restriction sites for insertion of a protein coding sequence. The constructed vectors placed the 3-hydroxysteroid oxidase gene under the control of the FMV35S promoter with the petunia HSP70 leader sequence discussed above. At the 3' end terminator
35 region is the non-translated polyadenylation signal terminator region of

the nopaline synthase gene. A plasmid containing the intact protein
coding sequence (SEQ ID NO:13) was identified and named pMON20910.
A plasmid containing the modified mature protein coding sequence (SEQ
ID NO:14) was identified and named pMON20908.

5   pMON20910 and pMON20908 are vectors for expression of
3-hydroxysteroid oxidase genes in plant cells, but these vectors lack
appropriate sequences for use in *Agrobacterium*-mediated plant transfor-
mation. However, these vectors can be used for either transient expres-
sion of 3-hydroxysteroid oxidase in plant cells, or they can be used to
10 generate stably transformed cotton plants via free DNA delivery such as
biolistic bombardment of cotton meristems.

       For transient expression analysis, plasmid DNA samples from
pMON20908 and pMON20910 vectors were purified and introduced into
tobacco via electroporation. Freeze-thaw extraction followed by a nine-
15 fold concentration of soluble fractions on Centricon-10 filter concentrators
allowed unambiguous detection of 3-hydroxysteroid oxidase activity in all
cell lysates, immunologically by Western blot assay and enzymatically.
The activity of the lysate from cells containing pMON20908, that is the
coding sequence for the modified mature protein, was approximately ten-
20 fold lower than that recovered from cells containing pMON20910.
Western blot analysis indicated that the signal sequence is cleaved in
protoplasts, although not necessarily with the fidelity necessary to
generate a processed protein identical in form and activity to that
naturally secreted by *Streptomyces* A19249.

25

Stable Transformation of Dicots with a 3-Hydroxysteroid Oxidase Gene

       pMON20908 and pMON20910 were used to construct vectors for
stable transformation of dicots with *Agrobacterium*. Each was partially
digested with the restriction enzyme NotI to generate DNA fragments
30 containing the FMV35S promoter, the petunia HSP70 leader, either the
intact [full length] or mature 3-hydroxysteroid oxidase coding sequence,
and nos 3' polyadenylation site. Partial digestion was necessary because
in addition to NotI sites flanking the expression cassette, there are two
NotI sites within the cholesterol oxidase gene - one within the coding region
35 for the signal peptide and one within the coding region for the mature pro-

tein. The NotI partial digest fragments containing the entire expression cassettes are isolated using agarose gel electrophoresis and purified by extraction from the agarose. These NotI fragments are inserted into NotI-digested plasmid pMON17227, which is described by Barry et al. in

5 WO 92/04449. pMON20913 was identified as containing the intact coding sequence. pMON20923 was identified as containing the mature coding sequence.

These vectors were introduced into disarmed *Agrobacterium* host ABI and used to transform tobacco in tissue culture. Selection for glypho-

10 sate resistance led to several transformed lines for each vector. For pMON20913, four out of 18 lines were confirmed as 3-hydroxysteroid oxidase expressors by Western blot assay. For pMON20923, four out of 29 lines were confirmed as 3-hydroxysteroid oxidase expressors by Western blot assay. Analysis of the lines transformed with pMON20913

15 confirm the presence of the 3-hydroxysteroid oxidase activity. The highest expressing plant had an expression level of 0.2% of total protein, which is equivalent to 54 ng of enzyme per mg of wet tissue.

Vectors containing the intact or mature 3-hydroxysteroid oxidase cassette express the active enzyme in the cytoplasm of the plant cell.

20 There has been no evidence of secretion outside the transformed cells. Some bacterial secretory signal sequences have been shown to function in plant cells. It may be desirable to direct most or all of the 3-hydroxy-steroid oxidase protein into the plant secretory pathway. To achieve this, it may be advantageous to use a signal sequence derived from a plant

25 gene rather than a bacterial signal. An example of such a signal is that from the tobacco PR1b gene, described by Cornelissen et al. pMON 10824, disclosed in EP Publ. 0 385 962, is a plant transformation vector designed for the expression of the lepidopteran active *B.t. kurstaki* protein. In pMON10824, the *B.t.k.* coding sequence is fused to the PR1b signal

30 sequence plus 10 amino acids of the mature PR1b coding sequence. This *B.t.k.* fusion gene is driven by the CaMV35S promoter containing a duplicated enhancer. Other vectors carrying the PR1B signal and CaMV35S promoter may also serve as the source of these elements. To create a vector in which the PR1b signal is fused to the 3-hydroxysteroid

35 oxidase gene, a DNA fragment containing the CaMV35S promoter

22

sequence and the PR1B signal sequence is used to replace the fragment
containing the FMV promoter and HSP70 leader sequences in pMON
20913 and pMON20923. This results in plasmids in which the 3-hydroxy-
steroid oxidase coding sequence (either mature protein or intact protein
5   cassette) is fused to the amino terminal secretory signal from the PR1b
gene and is driven by the CaMV35S promoter. NcoI restriction enzyme
sites at the 3' end of the PR1B signal-containing fragment and the 5' end
of the modified 3-hydroxysteroid oxidase protein coding sequence allows in-
frame translational fusions between the two elements. pMON20930 and
10   pMON20932 are plant transformation vectors that carry such fusions
between PR1B signal sequence and either the intact cholesterol oxidase
and the mature cholesterol oxidase, respectively. A similar plasmid may
be constructed wherein the 3-hydroxysteroid oxidase gene is under the
control of the FMV35S promoter. Such plasmids are mobilized into a
15   disarmed *Agrobacterium* host and used to transform dicots. Thus, these
plants produce a 3-hydroxysteroid oxidase that is secreted into the
extracellular space.

     In some cases, proteins that enter the plant secretory pathway are
targeted to different cellular compartments such as the vacuole. It may
20   be desirable to direct the 3-hydroxysteroid oxidase to the vacuole of plant
cells. In this case, the vectors described above in which the PR1b signal is
used are further modified to include vacuolar targeting sequences derived
from known plant vacuolar enzyme genes.

     It may also be desirable to retain the 3-hydroxysteroid oxidase in
25   the lumen of the endoplasmic reticulum (ER). In this case, vectors in
which the PR1b signal is used to target the protein to the secretory
pathway are further modified to include sequences known to encode ER
retention signals. These sequences are added such that the four-amino
acid-long ER retention peptides (such as HDEL (SEQ ID NO:15) and
30   KDEL (SEQ ID NO:16) are fused to the carboxy terminus of the 3-
hydroxysteroid oxidase. Site-directed oligonucleotide mutagenesis is used
to introduce these additions to the carboxy-terminus coding sequence of 3-
hydroxysteroid oxidase. pMON20937 and pMON20938 are vectors in
which the peptides HDEL (SEQ ID NO:15) and RGSEKDEL (SEQ ID
35   NO:17) are introduced, respectively, into pMON20932.

It may also be advantageous to direct the localization of the
3-hydroxysteroid oxidase protein to another cellular compartment, the
chloroplast. Proteins can be directed to the chloroplast by including at
their N-termini a chloroplast transit peptide (CTP). One CTP that has

5   worked to localize heterologous proteins to the chloroplast is that derived
from the RUBISCO small subunit gene of *Arabidopsis*, denoted ats1A. A
variant of this transit peptide that encodes the transit peptide, 24 amino
acids of mature RUBISCO sequence, plus a reiteration of the transit pep-
tide cleavage site has been constructed for the successful chloroplast

10 localization of the *B.t.k.* protein (Wong, 1992). Vectors containing the
Arabidopsis ats1A transit peptide fused to the *B.t.k.* gene may be used as
the base for constructing vectors for the chloroplast localization of the
3-hydroxysteroid oxidase protein. For example, pMON10817, constructed
of the *Arabidopsis* RUBISCO small subunit promoter from the *ats1A*

15 gene, the native *ats1A* 5' untranslated leader, the *ats1A* chloroplast
transit peptide variant described above and a truncated *B.t.k.* enzyme
gene is digested with restriction enzymes NotI and NcoI. A NotI - NcoI
DNA fragment containing the promoter, leader and transit peptide is used
to replace a NotI - NcoI fragment carrying the FMV promoter and HSP70

20 leader sequences in pMON20913 and pMON20923. These reactions
construct plasmids pMON20929 and pMON20931 in which the 3-hydroxy-
steroid oxidase coding sequence (either intact protein or mature protein
cassette, respectively) is fused at its amino terminus to the chloroplast
transit peptide and is transcribed from the *ats1A* promoter. Alternatively,

25 a similar plasmid may be constructed to replace the *ats1A* promoter with
the CaMV35S or FMV35S promoters. Such plasmids are mobilized into
disarmed *Agrobacterium* hosts and used to transform dicots. Thus, these
plants produce a 3-hydroxysteroid oxidase that is localized to the
chloroplast.

30   It may also be advantageous to direct the 3-hydroxysteroid oxidase
to another subcellular compartment, the mitochondria. Examples of
proteins that are normally localized to the mitochondria and for which the
targeting sequences are known include cytochrome C1 which is localized
to the mitochondrial intermembrane space (Braun, 1992) and the ß

35 subunit of ATP synthase which is localized to the mitochondrial matrix

(Boutry, 1987). Mitochondrial targeting DNA sequences encoding the amino terminal mitochondrial targeting peptides are isolated using the polymerase chain reaction (PCR). Oligonucleotide primers corresponding to the amino and carboxy terminal portions of these targeting sequences 5 are used to PCR-amplify first strand cDNA that has been generated from mRNA using reverse transcriptase. The oligonucleotide primers have NcoI restriction enzyme sites attached so that the amplified product, when cleaved with NcoI, can be inserted into the NcoI sites at the amino terminus of the cloned 3-hydroxysteroid oxidase gene cassettes. Thus, 10 various forms of 3-hydroxysteroid oxidase fused to mitochondrial targeting peptides can be expressed in plants using promoters such as CaMV35S or FMV35S, and can be localized to the mitochondria.

Stable Transformation of Monocots

15   A 3-hydroxysteroid oxidase gene may be stably incorporated into the genome of monocots using the vectors and methods described in WO 93/19189 (Brown et al.). The gene can be inserted in an appropriate vector, for example pMON19653 and pMON19643, described by Brown et al. The resulting construct contains a cassette of the CaMV E35S 20 promoter, the Hsp70 intron, the CP4 glyphosate selection marker, and the NOS terminator; a cassette of the CaMV E35S promoter, the Hsp70 intron, the GOX glyphosate selection marker, and the NOS terminator; and a single NotI site for insertion of a gene expression cassette containing a 3-hydroxysteroid oxidase gene, such as SEQ ID NO:13 or 25 SEQ ID NO:14.

        This vector is inserted by bombardment of embryogenic tissue culture cells using a biolistic particle gun as described by Brown et al. Transformed cells are selected for glyphosate resistance and whole plants are regenerated. Insect-resistant plants may be confirmed to be 30 expressing the gene by Western blot analysis, esterase activity assay, or insect resistance assay.

        Targeting of the protein to certain cellular compartments is also possible in monocots using the signal sequences described above.

        From the foregoing, it will be seen that this invention is one well 35 adapted to attain all the ends and objects hereinabove set forth together

with advantages which are obvious and which are inherent to the invention. It will be understood that certain features and subcombinations are of utility and may be employed without reference to other features and subcombinations. This is contemplated by and is within the scope of the
5 claims. Since many possible embodiments may be made of the invention without departing from the scope thereof, it is to be understood that all matter herein set forth or shown in the accompanying drawings is to be interpreted as illustrative and not in a limiting sense.

26

REFERENCES

Bagdasarian, M., et al. Gene, 16: 237-47, 1981.

Bevan, M. et al., Nature, 304:184, 1983.

Boutry, et al., Nature 328: 340-342, 1987.

5 Braun, et al., Mol. Gen. Genet. 231: 217-225, 1992.

Burnette, W.N. Anal. Biochem., 112: 195-203, 1981.

Cornelissen, B.J.C., et al. EMBO Journal, 5: 37-40, 1986.

Fischhoff, D.A. and Perlak, F.J. "Synthetic plant genes and method for preparation." European Patent Application, Publication Number 0

10   385 962, 1990.

Gallo, L.L. Methods Enzymol., 71: 665-7, 1981.

Herrera-Estrella, L. et al., Nature, 303:209, 1983.

Ishizaki, T., et al. Journal of Bacteriology, 171: 596-601, 1989.

Kay, R. et al., Science, 236: 1299-1302, 1987.

15 Klee, H. J. et al., Bio/Technology, 3: 637-642, 1985.

Knauf, V.C. and Nester, E. Plasmid, 8: 43-54, 1982.

Laemmli, U.K. Nature, 227: 680-5, 1970.

Long, Susan, and Ostroff, Gary R. PCT Int. Appl. WO 90 05,788.

Marrone, P.G., et al. J. Econ. Entom., 78: 290-3, 1985.

20 Matsudaira, P. J. Biol. Chem., 261: 10035-38, 1987.

Moore, S. and Stein, W.H. Methods in Enzymology, 6: 819-31, 1963.

Purcell, J.P., Greenplate, J.T., and Sammons, R.P. Insect Biochem. Molec. Biol., 22:41-47, 1992.

Purcell, J.P., et al. Biochem. Biophys. Res. Comm. 196:1406-1413, 1993.

25 Schuler, M. A. et al., Nucleic Acids Research, 10: 8225-8244, 1982.

Smith, A. G., and Brooks, C.J.W. J. Steroid Biochem., 7: 705-713, 1976.

Smith, P.K., et al. Analytical Biochemistry, 150: 76-85, 1985.

Winter et al. Mol. Gen. Genet., 221(2): 315-19, 1988.

Wong, et al., Plant Mol. Biol. 20:81-93, 1992.

30 Yanisch-Perron, C., et al. Gene, 33: 103-19, 1985.

27

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT:
      (A) NAME: Monsanto Company
      (B) STREET: 800 North Lindbergh Boulevard
      (C) CITY: St. Louis
      (D) STATE: Missouri
      (E) COUNTRY: United States of America
      (F) POSTAL CODE (ZIP): 63167
      (G) TELEPHONE: (314)694-3131
      (H) TELEFAX: (314)694-5435

    (ii) TITLE OF INVENTION: Method of Controlling Insects

    (iii) NUMBER OF SEQUENCES: 14

    (iv) COMPUTER READABLE FORM:
      (A) MEDIUM TYPE: Floppy disk
      (B) COMPUTER: IBM PC compatible
      (C) OPERATING SYSTEM: PC-DOS/MS-DOS
      (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

    (vi) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER: US 07/762682
      (B) FILING DATE: 23-SEP-1991

    (vi) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER: US 07/937195
      (B) FILING DATE: 09-SEP-1992

    (vi) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER: US 08/083948
      (B) FILING DATE: 28-JUN-1993

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 15 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Val Ser Thr Leu Met Leu Glu Met Gly Gln Leu Trp Asn Gln Pro
1   5   10   15

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 16 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Ala Phe Ala Asp Asp Phe Cys Tyr His Pro Leu Gly Gly Cys Val Leu
1   5   10   15

(2) INFORMATION FOR SEQ ID NO:3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 12 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Asn Leu Tyr Val Thr Asp Gly Ser Leu Ile Pro Gly
1   5   10

(2) INFORMATION FOR SEQ ID NO:4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 45 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTGTCCACCC TGATGCTGGA GATGGGCCAG CTGTGGAACC AGCCC   45

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 48 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCCTTCGCCG ACGACTTCTG CTACCACCCG CTCGGCGGCT GCGTCCTG   48

(2) INFORMATION FOR SEQ ID NO:6:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 36 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AACCTCTACG TGACCGACGG TTCGCTGATC CCGGGT   36

(2) INFORMATION FOR SEQ ID NO:7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1865 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CTACTCCATG GCGTGCTGAA GGTCGGTGCC TGGCCTCCCG AGGTCGTCGA GGACTTCGTG    60

AAGTGAGCGG GCACCCCGCC CGTCCCCGCC CCGCAACGGC CCGTTCCGCA CACCGGGTGA   120

CCCGACCCCC TCGGCCCCCG ACGTCCGCCG ACCTCTCAGT CCCCTCTCGA AGCTCAGGAG   180

CAACAGCGTG AACGCACACC AGCCTCTGTC GCGCCGCCGC ATGCTCGGCC TGGCCGCCTT   240

GGGCGCCGCC GCACTCACCG GCAGACCAC GATCACCGCG GCCCCCCGCG CGGCCGCCGC   300

CACCGCCCCC GGCGGCTCCG GCGGCACGTT CGTGCCCGCC GTCGTGATCG GCACCGGCTA   360

CGGCGCGGCC GTCTCCGCCC TGCGGCTCGG CGAGGCCGGG GTCTCCACCC TGATGCTGGA   420

GATGGGCCAG CTGTGGAACC AGCCCGGCCC GGACGGCAAC GTCTTCTGCG GGATGCTCAA   480

GCCCGACAAG CGCTCCAGCT GGTTCAAGAC CCGCACCGAG GCCCCGCTCG GCTCCTTCCT   540

CTGGCTCGAC CTCGCCAACC GGGACATCGA CCCCTACGCG GGCGTCCTGG ACCGGGTCAA   600

CTTCGACCAG ATGTCCGTGT ACGTGGGCCG CGGGGTCGGC GGCGGCTCGC TCGTCAACGG   660

CGGTATGGCC GTCACGCCCC GGCGCTCCTA CTTCCAGGAG GTGCTGCCCC AGGTCGACGC   720

CGACGAGATG TACGGCACCT ACTTCCCGCG CGCGAACTCC GGCCTGCGGG TCAACAACAT   780

CGACAAGGAC TGGTTCGAGC AGACCGAGTG GTACACGTTC GCGCGCGTTG CCCGTCTGCA   840

GGCCGAGAAC GCCGGCCTGA AGACCACCTT CGTGCCCAAC GTCTACGACT GGGACTACAT   900

GCGCGGTGAG GCGGACGGCA CCAACCCCAA GTCCGCGCTC GCCGCCGAGG TCATCTACGG   960

CAACAACCAC GGCAAGGTCT CCCTCGACAA GAGCTACCTG GCGGCCGCCC TGGGCACCGG  1020

CAAGGTCACC GTCGAGACCC TGCACCAGGT CAAGACGATC CGTCAGCAGA ACGACGGCAC  1080

CTACCTGCTG ACGGTCGAGC AGAAGGACCC CGACGGCAAG CTGCTCGGGA CCAAGGAGAT  1140

CTCCTGCCGC CACCTCTTCC TCGGCGCCGG CAGCCTCGGC TCCATTGAAC TGCTGCTGCG  1200

CGCCCGGGAG ACCGGCACCC TGCCCGGCCT CAGCTCCGAG ATCGGCGGCG GCTGGGGCCC  1260

CAACGGCAAC ATCATGACCG CCCGCGCCAA CCATGTGTGG AACCCCACGG GCAGCAAGCA  1320

GTCGTCGATC CCCGCCCTCG GCATCGACGA CTGGGACAAC CCCGACAACC CCGTCTTCGC  1380
```

CGAGATAGCC CCCATGCCGG CGGGCCTCGA GACCTGGGTC AGCCTCTACC TGGCCATCAC 1440

CAAGAACCCG GAGCGCGGCA CCTTCGTCTA CGACGCCGCC AAGGACCGGG CGGACCTGCG 1500

CTGGACCCGG GACCAGAACG CGCCCGCGGT CGCCGCCGCC AAGTCGCTGT TCGACCGCGT 1560

CAACAAGGCC AACACGACCA TCTACCGGTA CGACCTCTTC GGCAAGCAGA TCAAGGCGTT 1620

CGCCGACGAC TTCTGCTACC ACCCGCTCGG CGGCTGCGTC CTCGGCAAGG CCACCGACAA 1680

CTACGGCCGC GTCTCCGGGT ACAAGAACCT CTACGTCACC GACGGCTCGC TCATCCCCGG 1740

CAGCATCGGC GTCAACCCGT TCGTGACCAT CACGGCGCTG GCGGAGCGGA ACGTCGAGCG 1800

CGTCATCAAG GAGGACATCG CGGGTTCCTG ACGAGCGACG GGCGGGGCGC GGCATGCAAG 1860

CTTGG   1865

(2) INFORMATION FOR SEQ ID NO:8:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 547 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Asn Ala His Gln Pro Leu Ser Arg Arg Arg Met Leu Gly Leu Ala
1   5   10   15

Ala Leu Gly Ala Ala Ala Leu Thr Gly Gln Thr Thr Ile Thr Ala Ala
            20   25   30

Pro Arg Ala Ala Ala Ala Thr Ala Pro Gly Gly Ser Gly Gly Thr Phe
        35   40   45

Val Pro Ala Val Val Ile Gly Thr Gly Tyr Gly Ala Ala Val Ser Ala
        50   55   60

Leu Arg Leu Gly Glu Ala Gly Val Ser Thr Leu Met Leu Glu Met Gly
65   70   75   80

Gln Leu Trp Asn Gln Pro Gly Pro Asp Gly Asn Val Phe Cys Gly Met
            85   90   95

Leu Lys Pro Asp Lys Arg Ser Ser Trp Phe Lys Thr Arg Thr Glu Ala
        100   105   110

Pro Leu Gly Ser Phe Leu Trp Leu Asp Leu Ala Asn Arg Asp Ile Asp
        115   120   125

Pro Tyr Ala Gly Val Leu Asp Arg Val Asn Phe Asp Gln Met Ser Val
        130   135   140

Tyr Val Gly Arg Gly Val Gly Gly Gly Ser Leu Val Asn Gly Gly Met
145   150   155   160

Ala Val Thr Pro Arg Arg Ser Tyr Phe Gln Glu Val Leu Pro Gln Val

				165				170				175	

5

Asp Ala Asp Glu Met Tyr Gly Thr Tyr Phe Pro Arg Ala Asn Ser Gly
       180   185   190

Leu Arg Val Asn Asn Ile Asp Lys Asp Trp Phe Glu Gln Thr Glu Trp
    195   200   205

10

Tyr Thr Phe Ala Arg Val Ala Arg Leu Gln Ala Glu Asn Ala Gly Leu
    210   215   220

Lys Thr Thr Phe Val Pro Asn Val Tyr Asp Trp Asp Tyr Met Arg Gly
225   230   235   240

15

Glu Ala Asp Gly Thr Asn Pro Lys Ser Ala Leu Ala Ala Glu Val Ile
       245   250   255

Tyr Gly Asn Asn His Gly Lys Val Ser Leu Asp Lys Ser Tyr Leu Ala
       260   265   270

20

Ala Ala Leu Gly Thr Gly Lys Val Thr Val Glu Thr Leu His Gln Val
    275   280   285

25

Lys Thr Ile Arg Gln Gln Asn Asp Gly Thr Tyr Leu Leu Thr Val Glu
    290   295   300

Gln Lys Asp Pro Asp Gly Lys Leu Leu Gly Thr Lys Glu Ile Ser Cys
305   310   315   320

30

Arg His Leu Phe Leu Gly Ala Gly Ser Leu Gly Ser Ile Glu Leu Leu
       325   330   335

Leu Arg Ala Arg Glu Thr Gly Thr Leu Pro Gly Leu Ser Ser Glu Ile
       340   345   350

35

Gly Gly Gly Trp Gly Pro Asn Gly Asn Ile Met Thr Ala Arg Ala Asn
       355   360   365

40

His Val Trp Asn Pro Thr Gly Ser Lys Gln Ser Ser Ile Pro Ala Leu
    370   375   380

Gly Ile Asp Asp Trp Asp Asn Pro Asp Asn Pro Val Phe Ala Glu Ile
385   390   395   400

45

Ala Pro Met Pro Ala Gly Leu Glu Thr Trp Val Ser Leu Tyr Leu Ala
       405   410   415

Ile Thr Lys Asn Pro Glu Arg Gly Thr Phe Val Tyr Asp Ala Ala Lys
       420   425   430

50

Asp Arg Ala Asp Leu Arg Trp Thr Arg Asp Gln Asn Ala Pro Ala Val
       435   440   445

Ala Ala Ala Lys Ser Leu Phe Asp Arg Val Asn Lys Ala Asn Thr Thr
    450   455   460

55

Ile Tyr Arg Tyr Asp Leu Phe Gly Lys Gln Ile Lys Ala Phe Ala Asp
465   470   475   480

60

Asp Phe Cys Tyr His Pro Leu Gly Gly Cys Val Leu Gly Lys Ala Thr
       485   490   495

```
       Asp Asn Tyr Gly Arg Val Ser Gly Tyr Lys Asn Leu Tyr Val Thr Asp
               500             505             510

       Gly Ser Leu Ile Pro Gly Ser Ile Gly Val Asn Pro Phe Val Thr Ile
  5            515             520             525

       Thr Ala Leu Ala Glu Arg Asn Val Glu Arg Val Ile Lys Glu Asp Ile
               530             535             540

 10    Ala Gly Ser
       545
```

 (2) INFORMATION FOR SEQ ID NO:9:
 15
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 27 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
 20   (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
 25
 CTCAGGAGCA CCATGGCGAC CGCACAC   27

        (2) INFORMATION FOR SEQ ID NO:10:
 30
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 29 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
 35   (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
 40
  GTGCCGCCGG AGGCCATGGG GGCGGTGGC   29

        (2) INFORMATION FOR SEQ ID NO:11:
 45
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 30 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
 50   (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: DNA (genomic)

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
 55
 GCCCCGCCCG TCGGATCCGT CAGGAACCCG   30

        (2) INFORMATION FOR SEQ ID NO:12:
 60
        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 5 amino acids
```

33

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ser Gly Gly Thr Phe
1               5

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 1647 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
ATGGCGACCG CACACCAGCC TCTGTCGCGC CGCCGCATGC TCGGCCTGGC CGCCTTGGGC   60

GCCGCCGCAC TCACCGGGCA GACCACGATC ACCGCGGCCC CCCGCGCGGC CGCCGCCACC  120

GCCCCCGGCG GCTCCGGCGG CACGTTCGTG CCCGCCGTCG TGATCGGCAC CGGCTACGGC  180

GCGGCCGTCT CCGCCCTGCG GCTCGGCGAG GCCGGGGTCT CCACCCTGAT GCTGGAGATG  240

GGCCAGCTGT GGAACCAGCC CGGCCCGGAC GGCAACGTCT TCTGCGGGAT GCTCAAGCCC  300

GACAAGCGCT CCAGCTGGTT CAAGACCCGC ACCGAGGCCC CGCTCGGCTC CTTCCTCTGG  360

CTCGACCTCG CCAACCGGGA CATCGACCCC TACGCGGGCG TCCTGGACCG GGTCAACTTC  420

GACCAGATGT CCGTGTACGT GGGCCGCGGG GTCGGCGGCG CTCGCTCGT CAACGGCGGT  480

ATGGCCGTCA CGCCCCGGCG CTCCTACTTC CAGGAGGTGC TGCCCCAGGT CGACGCCGAC  540

GAGATGTACG GCACCTACTT CCCGCGCGCG AACTCCGGCC TGCGGGTCAA CAACATCGAC  600

AAGGACTGGT TCGAGCAGAC CGAGTGGTAC ACGTTCGCGC GCGTTGCCCG TCTGCAGGCC  660

GAGAACGCCG GCCTGAAGAC CACCTTCGTG CCCAACGTCT ACGACTGGGA CTACATGCGC  720

GGTGAGGCGG ACGGCACCAA CCCCAAGTCC GCGCTCGCCG CCGAGGTCAT CTACGGCAAC  780

AACCACGGCA AGGTCTCCCT CGACAAGAGC TACCTGGCGG CCGCCCTGGG CACCGGCAAG  840

GTCACCGTCG AGACCCTGCA CCAGGTCAAG ACGATCCGTC AGCAGAACGA CGGCACCTAC  900

CTGCTGACGG TCGAGCAGAA GGACCCCGAC GGCAAGCTGC TCGGGACCAA GGAGATCTCC  960

TGCCGCCACC TCTTCCTCGG CGCCGGCAGC CTCGGCTCCA TTGAACTGCT GCTGCGCGCC 1020

CGGGAGACCG GCACCCTGCC CGGCCTCAGC TCCGAGATCG GCGGCGGCTG GGGCCCCAAC 1080

GGCAACATCA TGACCGCCCG CGCCAACCAT GTGTGGAACC CCACGGGCAG CAAGCAGTCG 1140

TCGATCCCCG CCCTCGGCAT CGACGACTGG GACAACCCCG ACAACCCCGT CTTCGCCGAG 1200
```

```
ATAGCCCCCA TGCCGGCGGG CCTCGAGACC TGGGTCAGCC TCTACCTGGC CATCACCAAG 1260

AACCCGGAGC GCGGCACCTT CGTCTACGAC GCCGCCAAGG ACCGGGCGGA CCTGCGCTGG 1320

ACCCGGGACC AGAACGCGCC CGCGGTCGCC GCCGCCAAGT CGCTGTTCGA CCGCGTCAAC 1380

AAGGCCAACA CGACCATCTA CCGGTACGAC CTCTTCGGCA AGCAGATCAA GGCGTTCGCC 1440

GACGACTTCT GCTACCACCC GCTCGGCGGC TGCGTCCTCG GCAAGGCCAC CGACAACTAC 1500

GGCCGCGTCT CCGGGTACAA GAACCTCTAC GTCACCGACG GCTCGCTCAT CCCCGGCAGC 1560

ATCGGCGTCA ACCCGTTCGT GACCATCACG GCGCTGGCGG AGCGGAACGT CGAGCGCGTC 1620

ATCAAGGAGG ACATCGCGGG TTCCTGA                                      1647
```

(2) INFORMATION FOR SEQ ID NO:14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1521 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
ATGGCCTCCG GCGGCACGTT CGTGCCCGCC GTCGTGATCG GCACCGGCTA CGGCGCGGCC   60

GTCTCCGCCC TGCGGCTCGG CGAGGCCGGG GTCTCCACCC TGATGCTGGA GATGGGCCAG  120

CTGTGGAACC AGCCCGGCCC GGACGGCAAC GTCTTCTGCG GGATGCTCAA GCCCGACAAG  180

CGCTCCAGCT GGTTCAAGAC CCGCACCGAG GCCCCGCTCG GCTCCTTCCT CTGGCTCGAC  240

CTCGCCAACC GGGACATCGA CCCCTACGCG GGCGTCCTGG ACCGGGTCAA CTTCGACCAG  300

ATGTCCGTGT ACGTGGGCCG CGGGGTCGGC GGCGGCTCGC TCGTCAACGG CGGTATGGCC  360

GTCACGCCCC GGCGCTCCTA CTTCCAGGAG GTGCTGCCCC AGGTCGACGC CGACGAGATG  420

TACGGCACCT ACTTCCCGCG CGCGAACTCC GGCCTGCGGG TCAACAACAT CGACAAGGAC  480

TGGTTCGAGC AGACCGAGTG GTACACGTTC GCGCGCGTTG CCCGTCTGCA GGCCGAGAAC  540

GCCGGCCTGA AGACCACCTT CGTGCCCAAC GTCTACGACT GGGACTACAT GCGCGGTGAG  600

GCGGACGGCA CCAACCCCAA GTCCGCGCTC GCCGCCGAGG TCATCTACGG CAACAACCAC  660

GGCAAGGTCT CCCTCGACAA GAGCTACCTG GCGGCCGCCC TGGGCACCGG CAAGGTCACC  720

GTCGAGACCC TGCACCAGGT CAAGACGATC CGTCAGCAGA CGACGGCAC CTACCTGCTG  780

ACGGTCGAGC AGAAGGACCC CGACGGCAAG CTGCTCGGGA CCAAGGAGAT CTCCTGCCGC  840

CACCTCTTCC TCGGCGCCGG CAGCCTCGGC TCCATTGAAC TGCTGCTGCG CGCCCGGGAG  900

ACCGGCACCC TGCCCGGCCT CAGCTCCGAG ATCGGCGGCG CTGGGGCCC CAACGGCAAC  960
```

```
ATCATGACCG CCCGCGCCAA CCATGTGTGG AACCCCACGG GCAGCAAGCA GTCGTCGATC 1020

CCCGCCCTCG GCATCGACGA CTGGGACAAC CCCGACAACC CCGTCTTCGC CGAGATAGCC 1080

CCCATGCCGG CGGGCCTCGA GACCTGGGTC AGCCTCTACC TGGCCATCAC CAAGAACCCG 1140

GAGCGCGGCA CCTTCGTCTA CGACGCCGCC AAGGACCGGG CGGACCTGCG CTGGACCCGG 1200

GACCAGAACG CGCCCGCGGT CGCCGCCGCC AAGTCGCTGT TCGACCGCGT CAACAAGGCC 1260

AACACGACCA TCTACCGGTA CGACCTCTTC GGCAAGCAGA TCAAGGCGTT CGCCGACGAC 1320

TTCTGCTACC ACCCGCTCGG CGGCTGCGTC CTCGGCAAGG CCACCGACAA CTACGGCCGC 1380

GTCTCCGGGT ACAAGAACCT CTACGTCACC GACGGCTCGC TCATCCCCGG CAGCATCGGC 1440

GTCAACCCGT TCGTGACCAT CACGGCGCTG GCGGAGCGGA ACGTCGAGCG CGTCATCAAG 1500

GAGGACATCG CGGGTTCCTG A                                        1521
```

(2) INFORMATION FOR SEQ ID NO:15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

    His Asp Glu Leu
    1

(2) INFORMATION FOR SEQ ID NO:16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

    Lys Asp Glu Leu
    1

(2) INFORMATION FOR SEQ ID NO:17:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 8 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

    Arg Gly Ser Glu Lys Asp Glu Leu
    1   5

WHAT IS CLAIMED IS:

1.   A method of controlling lepidopteran insect infestation of plants comprising providing a 3-hydroxysteroid oxidase for ingestion by the insect.

5   2.   The method of Claim 1 wherein the insect is in a larval stage.

3.   The method of Claim 1 wherein said 3-hydroxysteroid oxidase is provided by plant-colonizing microorganisms which produce 3-hydroxysteroid oxidase after application to the plant.

4.   The method of Claim 1 wherein said 3-hydroxysteroid oxidase is

10 provided by expression of a gene for 3-hydroxysteroid oxidase incorporated in the plant by previous genetic transformation of a parent cell of the plant.

5.   The method of Claim 4 wherein said plant is cotton or corn.

6.   A method of producing a genetically transformed plant which

15 expresses an amount of a 3-hydroxysteroid oxidase effective to control lepidopteran insect infestation, comprising the steps of:

        a)   inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising

                (i)   a promoter which functions in plant cells to cause the

20   production of an RNA sequence;

                (ii)   a structural coding sequence that encodes for 3-hydroxysteroid oxidase;

                (iii)   a 3' non-translated region which functions in said plant cells to cause the addition of polyadenylate nucleotides to the

25   3' end of the RNA sequence,

        wherein said promoter is heterologous with respect to said structural coding sequence and wherein said promoter is operatively linked with said structural coding sequence, which is in turn operably linked with said non-translated region;

30   b)   obtaining transformed plant cells; and

        c)   regenerating from the transformed plant cells genetically transformed plants with express an amount of 3-hydroxysteroid oxidase effective to control lepidopteran insect infestation.

7.   The method of Claim 6 wherein said structural DNA sequence

35 comprises SEQ ID NO:13 or SEQ ID NO: 14.

37

8.	The method of Claim 6 wherein said plant cell is a cotton or corn plant cell.

9.	The method of Claim 6 wherein wherein the genome of said plant cell also contains one or more genes expressing *B.t.* endotoxins.

```
ID   AF263912    standard; DNA; PRO; 123580 BP.
XX
AC   AF263912;
XX
SV   AF263912.1
XX
DT   25-MAY-2000 (Rel. 63, Created)
DT   25-MAY-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
DE   complete sequence.
XX
KW   .
XX
OS   Streptomyces noursei
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
XX
RN   [1]
RP   1-123580
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   "Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces
RT   noursei ATCC 11455: analysis of the gene cluster and deduction of the
RT   biosynthetic pathway";
RL   Chem. Biol. 7(6):395-403(2000).
XX
RN   [2]
RP   1-123580
RA   Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA   Valla S., Zotchev S.B.;
RT   ;
RL   Submitted (04-MAY-2000) to the RL   Unigen, NTNU, O. Kyrres gt. 3, Trondheim N-
XX
DR   SPTREMBL; Q9L4V6; Q9L4V6.
DR   SPTREMBL; Q9L4V7; Q9L4V7.
DR   SPTREMBL; Q9L4V8; Q9L4V8.
DR   SPTREMBL; Q9L4V9; Q9L4V9.
DR   SPTREMBL; Q9L4W0; Q9L4W0.
DR   SPTREMBL; Q9L4W1; Q9L4W1.
DR   SPTREMBL; Q9L4W2; Q9L4W2.
DR   SPTREMBL; Q9L4W3; Q9L4W3.
DR   SPTREMBL; Q9L4W4; Q9L4W4.
DR   SPTREMBL; Q9L4W5; Q9L4W5.
DR   SPTREMBL; Q9L4W6; Q9L4W6.
DR   SPTREMBL; Q9L4W7; Q9L4W7.
DR   SPTREMBL; Q9L4W8; Q9L4W8.
DR   SPTREMBL; Q9L4W9; Q9L4W9.
DR   SPTREMBL; Q9L4X0; Q9L4X0.
DR   SPTREMBL; Q9L4X1; Q9L4X1.
DR   SPTREMBL; Q9L4X2; Q9L4X2.
DR   SPTREMBL; Q9L4X3; Q9L4X3.
DR   SPTREMBL; Q9L4X4; Q9L4X4.
DR   SPTREMBL; Q9L4X5; Q9L4X5.
DR   SPTREMBL; Q9L4X6; Q9L4X6.
DR   SPTREMBL; Q9L4X7; Q9L4X7.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..123580
```

```
FT                      /db_xref="taxon:1971"
FT                      /organism="Streptomyces noursei"
FT                      /strain="ATCC 11455"
FT        CDS           complement(46..783)
FT                      /codon_start=1
FT                      /db_xref="SPTREMBL:Q9L4X7"
FT                      /note="putative 4'-phosphopantheteine transferase"
FT                      /transl_table=11
FT                      /gene="nysF"
FT                      /function="presumably post-translationally modifies ACP
FT                      domains on PKS"
FT                      /product="NysF"
FT                      /protein_id="AAF71762.1"
FT                      /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQREF
FT                      TTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSAAV
FT                      SIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYPLTQ
FT                      RELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHLPAPT
FT                      PRRDREHRTHLTVNSPLPRPTFG"
FT        CDS           complement(867..2684)
FT                      /codon_start=1
FT                      /db_xref="SPTREMBL:Q9L4X6"
FT                      /note="putative transporter (ABC family)"
FT                      /transl_table=11
FT                      /gene="nysG"
FT                      /function="presumably involved in efflux of nystatin"
FT                      /product="NysG"
FT                      /protein_id="AAF71763.1"
FT                      /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVLN
FT                      AFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGRLV
FT                      ASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAELIT
FT                      SIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTLNAHV
FT                      EEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANLGYVLV
FT                      AVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVFTLLDAP
FT                      EQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVGPTGAGKT
FT                      TLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIAENIAYGAP
FT                      GACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVARAFLARPAVL
FT                      VLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVMDAGRIVEQGT
FT                      HDQLLCAQGLYARLHAARTHTPTAGAAAG"
FT        CDS           complement(2662..4416)
FT                      /codon_start=1
FT                      /db_xref="SPTREMBL:Q9L4X5"
FT                      /note="putative transporter (ABC family)"
FT                      /transl_table=11
FT                      /gene="nysH"
FT                      /function="presumably involved in efflux of nystatin"
FT                      /product="NysH"
FT                      /protein_id="AAF71764.1"
FT                      /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQG
FT                      VVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFSAR
FT                      EIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALLLVA
FT                      LVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERARFAQ
FT                      TNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSALLSYL
FT                      SLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRIELCAAG
FT                      YRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAVRVGGTDV
FT                      RELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQAADFVARMP
FT                      DGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDAALRTALVPY
FT                      TAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREIALSQLTEEEA
FT                      AHGLAGRP"
FT        CDS           4714..5748
FT                      /codon_start=1
```

```
FT                        /db_xref="SPTREMBL:Q9L4X4"
FT                        /note="putative dGDP-mannose-4,6-dehydratase"
FT                        /transl_table=11
FT                        /gene="nysDIII"
FT                        /function="presumably involved in mycosamine biosynthesis"
FT                        /product="NysDIII"
FT                        /protein_id="AAF71765.1"
FT                        /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRVS
FT          RLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMGVL
FT          RMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYGAAK
FT          AYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKLALGN
FT          LDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNWEDYVV
FT          IDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQYGDVLL
FT          AANW"
FT     CDS                5930..34363
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4X3"
FT                        /note="polyketide synthase"
FT                        /transl_table=11
FT                        /gene="nysI"
FT                        /function="responsible for condensation steps 9 to 14 in
FT          the nystatin polyketide backbone synthesis"
FT                        /product="NysI"
FT                        /protein_id="AAF71766.1"
FT                        /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCRY
FT          PGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFFGI
FT          SPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADGAEG
FT          FQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLALAGGV
FT          TIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRNGHEIL
FT          AVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGTTLGDPV
FT          EAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTLPRTLHAE
FT          EPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAPEPQPEDAA
FT          TAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRPDLLDLAHSL
FT          ATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAETRLAVLFTGQG
FT          AQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALLDRTAYAQPALF
FT          AVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVAARGRLMQALPDG
FT          GAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTALAEQLAADGRKTR
FT          RLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPATVAQLTSADYWVDH
FT          VRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADADAVTLPALRAGRPEE
FT          HTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRYWPKALQSGTADLRSVG
FT          LGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRGTTLLPGTAFLELAVRAG
FT          DEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRTVNVHARPDTGDDTPWTAH
FT          ATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPPAGAEPLPLDGHYDRLADGG
FT          FGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGLHPALLDAALHAAPFTGLGER
FT          GRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTVADGTGAPVLSVDSLVLRSVAT
FT          QQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGADPFGLLTHAGFADAPAYPDLAA
FT          LAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAEALAAVQTWLDHHERFAAARLVFV
FT          TRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDLDPDGAVGAAALVAALVSGEPQLAV
FT          RGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSGVSFSGEGAVLVTGGTGGLGAVLARH
FT          LVAEYGVRDLLLVSRSGERAVGAGELVAELAGVGARVRVVACDVTDRAAVVELVGGHAV
FT          SAVVHAAGVLDDGMVGALTGERLSAVLRPKVDAVWHLHEATRGLDLDAFVVFSSLAGVF
FT          GSPGQANYAAANAFLDALMTRRRAEGLPGLSLAWGPWEQSGGMTGTLTDVDAERLARSG
FT          VPPLSVAQGLALFDAAVAGTDATCVPVRLDLPVLRARGEVPPLLRSLIRVARRAAVAG
FT          SATAGNLAQRLRRLDEDGRDEMVLDLVRGQVALVLGHATGGDVDAGRAFRDLGFDSLTA
FT          VELRNRLNTVTGLRLPATLVFDYPTVRHLATYVLDELLGTDAEVATVQPAAVAVADDPI
FT          VIVGMACRYPGGVSSPEDLWRVLTEGTDAVSGFPTNRGWDVESLYHPDPDHPGTSYTRS
FT          GGFLHEAGEFDPGFFGMSPREALATDSQQRLLLESSWEAIERAGIDPVSLRGSRTGVFA
FT          GVMYSDYSAMLASPEFEGFQGSGSSPSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHW
FT          AMQALRSGECGLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVG
FT          VLVLERQSDAVRNGHEILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAG
```

FT	DVDVVEAHGTGTTLGDPIEAQALLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVI
FT	KMVLAMRHGVVPRTLHVDAPSSHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGT
FT	NVHVIVEQAPGAKAIAAAGAARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAEL
FT	VDVALSLATTRSRFEQRAAVVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLF
FT	TGQGSQRAAMGRELHEVQPEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDET
FT	GWTQPALFAVEVALFRLVESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARAT
FT	LMQALPTGGAMIAIQATEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAER
FT	GRKTKRLRVSHAFHSPRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAE
FT	YWVRHVREAVRFADCVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAG
FT	RAEELAAATALARLQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALA
FT	AADPADQQLWAAVERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLR
FT	YRVEWTRLSKPTAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSC
FT	ADRAVLAERLARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQ
FT	GRLWTATRGAVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQR
FT	LAGILAVKDAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLG
FT	AEVARWLARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVL
FT	TALPEELPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFV
FT	LFGSIAGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSE
FT	TLRRQGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEV
FT	RALAAADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGR
FT	AFRDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGP
FT	VATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
FT	PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRPE
FT	TLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVDTA
FT	CSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFSDDA
FT	DGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQRVIRQ .
FT	ALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKSNIGHT ˙
FT	QSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPETGRPRRA ·
FT	AVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAAALRGQAE˙
FT	RLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALRAVATDGPS
FT	PVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAALAEFTDWSLV
FT	DVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQGEIAAAVVSGA
FT	LSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRVSVAAVNGPRSV
FT	VVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEVLAELAPRTSEVP
FT	FFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFVEVSSHPVLSMAVQ
FT	EAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGLFEGTGASRIDLPTY
FT	AFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALGTDEDSVAAVLPALTS
FT	WRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADGIDDTDVAGALETYGAE
FT	VRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAVPGTSLVLGTALTVALIQ
FT	ALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWTAALEHPQRWGGTLDLPAA
FT	LDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEATAGRPAGTWTPRGTTLVTG
FT	GSGTLAPHLARWLAQRGAEHLVLISRRGTAAPGAAELVAELAESGTEATVAACDITDRD
FT	AVAALLADLKADGRTVRTVVHTAATIELHTLDATTLADFDRVLHAKVTGAQVLAELLDD
FT	EELDDFVLYSSTAGMWGSGAHAAYVAGNAYLAALAEHRRANGLPALSLSWGIWADDLKL
FT	GRVDPQMIRRSGLEFMDPQLALSGLQRALDDNENVLAVADVDWETYHPVYTSGRPTPLF
FT	DEVPEVRRLTAAAEQSAGTVAEGEFAAALRALSDAEQQRTLLETVRTEAASVLGLSSAE
FT	DLTDQRAFRDVGFDSLTAVGLRNRLASVTGLTLPSTMVFDYPNPAALAAYLHGELAGAR
FT	SAAAGAAAVPTGAPDADDPIAIVGMSCRYPGGVGSAEDLWRIALDEVDAISGFPADRGW
FT	DAEGLYDPDPDRPGRTYSVQGGFLRDVAEFDPGFFGISPREALSMDPQQRLLLETAWEA
FT	FEHAGIDPVGQRGSRTGTFVGASYQDYASGVPNSEGSEGHMITGTLSSVLSGRVSYLFG
FT	FEGPAVTLDTACSSSLVAMHLACQSLRNGESSLALAGGVSIMSTPMSFVGFSRQRALAE
FT	DGRCKAYADGADGMTLAEGVGLVLLERLSDARANGHQVLAVIRGSAVNQDGASNGLTAP
FT	NGPSQQRVIRQALANSAVAPGDIDVLEGHGTGTALGDPIEAQALLATYGQDRAPERPLL
FT	LGSVKSNIGHTQMASGVASVIKLVRALQEGVVPKSLHIDRPSTHVDWSSGAIGLLTERT
FT	PWPETGRPRRAAVSSFGISGTNVHTILEQAPADEAPTPADPPRDGLVPVLLSGRGEAAL
FT	RAQAARLLAFVEERPEAHLTDLAHSLATSRAALERRAAVIAADRDTLTRGLRALSDGRP
FT	DPGLVQGTAGRGRTAFLFTGQGSQRPGMGRELHDRYPVFADALDEVLARLDDGPDRPLR
FT	EVLFAAPDSAEAALLDRTGYAQPALFAVEVALFRLLTSWGLTPDYLAGHSVGELAAAHV
FT	AGVLSLDDACTLVAARGRLMQALPEGGAMVALEAAEDEVLPLLEGLTDRVSVAAVNGPR
FT	SVVVAGVEEDVLLLADLFAADGRRTKRLRVSHAFHSPLMDAMLDDFAAVARGLTYHPPT

FT	IPFVSNVSGGLATAEQVRTPDYWVGHVRAAVRFADGIDWLATQGDVHTFLELGPDGVLS
FT	AMARESLTDPSRTALLPTLRGDRPEEPALVTAVAAAHAGARVDWSGYFADHGARRTTL
FT	PTYAFQRERYWPDTTAATSAHTPGSALDAEFWAAVERDDVAALAASLDLDDATVTAMVP
FT	ALTAWRRRGEQTELDSWRYRVTWKPRGGATAPAALTGRWLVLVPHDHQDRQDDATAAW
FT	AADVETALGTTTVRLTVTTTDRAALAARITEAAGDQGPFSGVLSLLPLATGDAGHPGAP
FT	AALTLTTTAVQALGDAGIDAPLWNVTRGAVAVGRAEQVTAPEQAAVWGLGRAVALELPA
FT	RFGGTLDLPATLDGQAARRLRAVLAATDGEDAVALRPSGVFLRRLAHAPAGPDTARTAF
FT	DPAAGTVLITGGTGGIGGHVARRLARDGATHLLLTSRRGPAAPGADALRAELEELGARV
FT	TLAACDAADRDALAALLAELPDDAPLCAVFHTAGVVEDHVVDALTPENFAAVLRAKTVA
FT	AHHLHELTADLDLAAFVLFSSTAGVLGAAGQGNYAAANAHLDALAEHRRSHGLTALSVA
FT	WGPWAGSGMVADAAELTDRVRRGGFEPLAPEPAVRALLRAIENDDTTVALADIDWERFQ
FT	RAFAAVRPLPFVADLPETGRATPATATGAATGLRQQLAELPEHERPAAVLDLLRTQVAA
FT	VLGHADPRTVEDDHAFRDLGFDSLTILELRNALNAATGLSLPATLVYDLPTPREMADFL
FT	LAELLGTLPTDTAATVASTASPKLSASFEQGGTPFDDPIAVIGIGCRFPGGVTTPEELW
FT	QLLDEGRDGISRFPDDRGWDLAALGAGASDTLEGGFLTGVADFDARFFGISPREALAMD
FT	PQQRLLLETTWEALERAGIDPTTLRGSTTGVFVGTNGQDYPTLLRRSASDVAGYVATGN
FT	TASVMSGRLSYALGLEGPAVTIDTACSSSLVALHWAGRALRAGECDLVVAGGVSVMASP
FT	DSFVEFSTQGGLAPDGRCKAFSDAADGTAWSEGVGILVLERLSAARRNGHQVLGLIRGT
FT	AVNQDGASNGLTAPNGLSQQRVIAQALADARLRPADIDAIEAHGTGTTLGDPIEARALI
FT	TAYGRDRDAERPLLLGTVKSNIGHTQAAAGAAGVIKMLMAMRHGTLPRTLHVGTPSSHV
FT	DWSGGTVALLDDARPWPRTGQPRRAGVSAFGVSGTNAHVVVEQAPETEAPAAPAAEPAP
FT	EATPTVVPWVVSGRSREALQAQLDRLTAHTAAHPARSAADVGRSLATDRTLFPHRAVLL
FT	AGPDGVREAARAAAPRTPGRTAFLFSGQGAQHALMGHDLYQRFPVYADALDTVLAQFDT
FT	VLDVPLRAALFAAPGTPEAALLDQTGFTQPALFAVEVALFRLAESWRLTPDFVAGHSIG
FT	EIAAAHVAGVFSLEDACTLVAARASLMQQLPRDGAMVALEATEDEVAPLLTDGVALAAV
FT	NGPRSVVVAGAEDAVRAVADRLAADGRRTRRLTVSHAFHSPLMDPMLTDFARVAEGLTY
FT	HEPRIPLVSTLLGAPAGAELRTPDYWVRHVRETVRFADGVRALHDAGAGTFVEIGPDGV
FT	LTALTQQTLDTVEAGAPAVVVPLQRRDRAGDLALLEGLATLHTHGTGPSWPAYFEATGG
FT	HRTDLPTYAFQRERYWPELGAPVATAPQDPAAWRYHETWAPLPAPEAAAPAGRALVLVP
FT	AGNRDTAWMTAVADALGADTVTAEPDALAEQLTAAGDTPWRVVVSLLAAASEGLPADGA
FT	WPAALLATLDEAGVHAPLWCVTRGAVAVAGEAPTAVGQAALWGLGRVAALDHPDRFGGL
FT	ADLPADTDAHAAGLLAAHLAAPGTEAEIAVRATGVHARRLVRTPAAADGATWLPTGTVL
FT	VVGGTGGTGTMGGRAARWLVREGARHLVLTAPDGTTTAADTEALTAELAALGARITVVD
FT	HDPTAPDGFAALLDGLPDDTPLTAVVYAPEADAAPGTAAELSAALAPVTALGAALTGRP
FT	LDAFVLFGSIAGLWGVRGRAAEAASGAYLDAFARACRDRGTPALAVAWGAWADLVGPSL
FT	AAHLRMNGLPVMDADTALTALSRAVADGSAAEAVADVRWETFAPLHHEARRTALFDALP
FT	EARGALAEAARDRADRKTAAGDYGRWLAEQPAADHDAILLALVTEKAATVLGHADHDLL
FT	EPDLPFRDLGFDSLTAVDLRNQLTAETGLTLPATLVFDHPNPAALAAHLRAQLLGEASD
FT	SAAPVAAPVALGADDDAIVIVGMACRYPGGVTSPEDLWQLVGDEVDAVGDFPTDRGWDL
FT	AALAGDGPGRSATAQGGFLYDATDFDPGLFGISPREALVMDPQQRILLETSWEALERAG
FT	IDPATLRGSGTTGVFVGGGSGDYRPPEEAGQWQTAQSASLLSGRLAYTFGIQGPTVSVD
FT	TACSSSLVALHLAAQALRAGECSIALAGGVTVMATPVGFVEFSAQGALSPDGRCRAFSD
FT	DANGTGWSEGVGMLVVERLSDARRNGHRVLAVLRGSAINQDGASNGLTAPSGPAQQRVI
FT	RQALANARLRPADIDAVEAHGTGTRLGDPIEAQALLATYGQDRERPVLLGSLKSNIGHT
FT	QAASGVGGVIKMVLAMQHGELPRSLYAENPSSHVDWTAGRAHLLTARTPWPDSGRPRRA
FT	AVSSFGASGTNAHAILEQPPREELPARPADDGAPLPFLLSGRSQNALRAQARRLLARLT
FT	AHPDTRAADLAYSLATTRAAFEHRAAITATDHDGLRTGLTAVAEGTTAPHTAEHHLQGT
FT	GKRAVLFSGQGSQRLGMGRELHERHPVFAEAFDSVLARLDDRLDTPLRDVVWGTDEEAL
FT	HATGNTQPALFAVEVALYRLIESWGVRPDFVAGHSVGELAAAHVAGVLSLDDACRLVAA
FT	RAALMQRLPAGGAMIAVEATEDEVTPLLTDGVSLAAVNGPTAVVLSGAGDAVTALGQAL
FT	AERGHRTTRLRVSHAFHSHLMDPMLADFRTVAEGLEYHPPRIPVVSNLTGDVADAADLC
FT	SADYWVRHVRGTVRFADGVRTMADRGVHLFLELGPDAVLSAMARQCAPDAVVVPALRRN
FT	RDEDETLVGAVARLHVHGAGPRWDAYFAGRGAQWLDLPTYPFQRGRFWPESLPGAASAA
FT	PAAGQPAETDAAFWDAVAQEDFTALESVLDVESDALSKVLPALMDWRSRQADESQLAGW
FT	RHRIVWKRLTGAALAHRKALSGTWLAVVPEGFADDPWVTTTLDGLGTHLVHLEVAEADR
FT	AALADAIAARTADGTRFGGVISLLALREELTGAVPEGTALTTTLLQALGDAGVDAPLWC
FT	VTRSAVSAGRTDRPHRPLQGAVWGLGRVAALEYPQRWGGLVDLPEEPDERSAAGLAAVL
FT	AGLDGEDQVAVRGTAVLARRLVPAPGRKPSRPWHPSGTVLVTGGTGALGAHVARRLAKD
FT	GAQHLVLLSRRGPDAPGAAELRAELDALGTDVTVAACDVADRDQLTAVLDALPADRPLT
FT	GVVHTAGVLDDGVLDRLTPERFQEVFRAKVTSALLLDELTRDRELAAFVLFSSASAAVG

```
FT                    NPGQANYAAANAVLDALAEQRRVLGLPATSVSWGAWGGGGMADADGADEAARRAGVGAM
FT                    DPHLAVEALLRLVAEKEPTAVVAEVALDRFAGAFGGSRPSALLREFPGYREALAAQAEQ
FT                    AADGGGLAARLAALPPARRLDTVVDLVRTRAAQVLGYPDTEAVAAERSFRDLGVDSLGA
FT                    VELRNQLSAATGLNLPATLVFDHPTPLVLGEHILGGLFPDEPAGSDDETEIRALLASVP
FT                    LDQLREIGVLEPLLQLAGRGGRAADGDDGESVDSMTVADLVRAALNGQSDL"
FT        CDS        34384..50691
FT                    /codon_start=1
FT                    /db_xref="SPTREMBL:Q9L4X2"
FT                    /note="polyketide synthase"
FT                    /transl_table=11
FT                    /gene="nysJ"
FT                    /function="responsible for condensation steps 15 to 17 in
FT                    the nystatin polyketide backbone synthesis"
FT                    /product="NysJ"
FT                    /protein_id="AAF71767.1"
FT                    /translation="MNAPENPETPENNVVAALRAAVKETDRLRRQNRMLVAAAKEPIAV
FT                    VGMACRFPGAVDSPEALWEMVATGTDVISGFPDDRGWDLEALRNSGTDARDTDVSQRGG
FT                    FLDCIADFDPGFFGISPREAVTMDPQQRLLLTTAWEAVERAGIDATTLRATRTGAFIGT
FT                    NGQDYAYLLVRSLDDATGDVGTGIAASAASGRLSYTLGLEGPALTVDTACSSSLVALHL
FT                    AVQALRNGECGMALAGGVNVMATPGSLVEFSRQGGLARDGRCKAFADAADGTGWSEGAG
FT                    VLLLERLSDAQRNGHPVLAVVRGSAVNQDGASNGFTAPNGPSQQRVIRQALANAGLATG
FT                    DIDAVEAHGTGTPLGDPIEAQSILATYGQDRAHPVLLGSIKSNMGHTQAASGVAGVIKM
FT                    IMAMRHGVLPRTLHVDRPSTHVDWTTGSVELLTDAHPWPETGRPRRTGISSFGVSGTNA
FT                    HVIVEQAPDTPAEAADDTPPRTPRTLPWLLSARTGAALRDQATALLDHLDRPDGDRGPT
FT                    ALDTAFSLATTRAALEHRLAVVTGTDGTAGRDALTAWLAHGTAPDAHEGHAAGRTRCAA
FT                    LFSGQGAQRLGMGRELHARFPVFARALDTAVDLLDAELGGTLREVIWGTDDAPLNETGF
FT                    TQPALFAVEVALYRLIESWGVAPDFVAGHSIGEIAAAHVAGVFSLEDACTLVAARAGLM
FT                    QALPRGGAMVAVEATEDEVSPLLTDGVAIAAINGPTSLVVSGDETATLAVAARLAEQGR
FT                    RTTRLRVSHAFHSPLMDPMLAEFRAVAEGLSYGEPQIPVVSNLTGAVADGTLLGTADYW
FT                    VRHVREAVRFADGIRALTDAGVGAFLELGPDGTLAALAQQSAPDAVSVPVLRKDRDEEP
FT                    AAVAALARLHTAGVPVDWTAFYAGTGAHRTDLPTYAFQYERYWPKATYRPADATGLGLT
FT                    AADHPLLGAAMSVAGSDELLLTGTLSLATHPWLADHVVGGMVFFPGTGFLELAVRAADQ
FT                    VGCDRVEELMLAAPLILPATGTVQMQIAVGAADDDGGRDLRFFTRPGDDPDAAWAQHAT
FT                    GRITEGERVLALDTTTWPPRDAEPVDIDGLYDRYRANGLDYGPVFRGLRAVWRRDTEIY
FT                    AEVALPEGTADADAFGLHPALFDAVLHSTLFASADGDDRSLLPFAWNGVSLHAAGADAL
FT                    RVRITSCGPDAVEITAVDPQGRPVVSVESLTLRAAGPDAGTADHRADAGSLFRMDWTPR
FT                    TVHAPATPATWAVLGTDPIGLTEALTAAGPDTVTGLRDGVDALGELTAGDDRPVPDVVA
FT                    VPLRGATDHGPAGAHDLTRTVLALLQEWLAEERFARSRLLLVTRGAVADGERGPLDLAA
FT                    APVWGLVRSAQSENPGRLLLVDLDDTAESAAQLPLLPALLDADEPQAVVREGTVRVGRL
FT                    ARLDSGRGLVPPPGTPWRLGSRAKGSLDGLALLPHPEARRPLTGHEVRVGIRAAGLNFR
FT                    DVLNALGMYPGDAGLFGSEAAGVVVEVGPEVTGLAPGDRVMGMLFGGFGPLGIADARLL
FT                    TPVPADWSWETGASVPLVFLTAYYALKELGGLRAGEKVLVHAGAGGVGMAAIQIARHVG
FT                    AEVFATASEGKWDVLRSLGVADDHIASSRTLDFEAAFAEVAGDRGLDVVLNALSGEFVD
FT                    ASMRLLGDGGRFLEMGKTDIRAADSVPDGLSYHSFDLGMVDPEHIQRMLLDLVELFDRG
FT                    ALAALPVRSWDVRRAGEAFRFMSLAQHIGKIVLTVPQPLDPDGTVLLTGGTGGLAGLLA
FT                    RHLVTEHGARHLLLAGRRGPDAPGAAALHAELTALGAEVTVAACDVADRTALAALLATV
FT                    PAEHPLTAVVHTAGVLDDGTLTALNPDRLATVLRPKVDAAWHLHDLTRHLDLAAFVLYS
FT                    STAGVMGGPGQANYAAGNTFLDALAAHRHALGLPATSLAWGAWEQGAGMTGALTDHDLR
FT                    RVSDAGGQPLLTAERGLALYDAATAADEPLIVPLGLTGGALPAGVGVPAVLRGLVRTAG
FT                    RRARAGTAGVSRAGLAERLAALPEEERTPFLVELVRTEAATVLGHGSTDPVDARREFRQ
FT                    LGFDSLTAIELRNRLGKATGLTLPATLIFDYPTPDRLAVHLHDELLGADAPVTVTAAAQ
FT                    AADPEHDPVVIVGMSCRFPGGVSSPEELWDLVASGTDAITGFPADRAWDRHPQLAGAPG
FT                    ARTGQGGFLRDIADFDAAFFGISPREALAMDPQQRILLEVAWEAAERAGIDPQTLRGSD
FT                    TGVFMGVSGQDYAGLVMRSRDDIAGHATTGLAVSVVSGRLAYALGLEGPALSVDTACSS
FT                    SLVSLHLAAQALRAGECTMALAGGVTVMTTAANFTGFSRMGGLAQDGRCKAFSDSADGT
FT                    GWSEGAAVLVLERLSDARRAGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALA
FT                    NAGLTPVDVDAVEAHGTGTPLGDPIEAQALIAAYGTDRDPEHPLLLGSVKSNIGHTQSA
FT                    AGAAGLVKMVMAMRHGILPQTLHLTEPSSHVDWSAGTVRLLTERTAWPRTDRPRRAGVS
FT                    SFGISGTNAHVILEQPPAEPTPAADPGRPAPTVVAWPVSAQTPAALDAQLDRLRTAAAL
FT                    APLDTAHTLATGRSLFEHRAVLLATVGDPATGAPDLPEVARGAATPHRTAFLFSGQGAQ
```

```
FT                       RSGMGRELHAAFPVFAAAFDEVVAVLDAELGSDADGGVSLREVMWGGGSELLDRTRFTQ
FT                       PALFAVEVALFRLVASWGVGPEFVAGHSVGEIAAAHVAGVFSLVDACRLVVARASLMDA
FT                      .LPVGGVMVAVEAAEAEVVPLLVDGVAIAAVNGPVSVVVSGVEAAVGQVVDQLVERGRRV
FT                       RRLAVSHAFHSPLMDPMLDAFRAVAEGLEYHQPRIPVVSNVTGEVAAAEELCAADYWVR
FT                       HVRATVRFADGVRTLAERGATAFLEIGPDGVLSALARGVLPAEALVTPTLRKDRDEESA
FT                       LLAGLARLHVAGVTVDWSAALTGTGARGTDLPTYAFQRERYWPELAAEPAGGGADAADA
FT                       EFWAAVERADATALAAHLDIDGDQLGAVLPALSAWRTRRRTTSATNALRHRESWEPLSL
FT                       AGTPHTGGVLVLVPAAATTDPWVADVVAALGPDARRVDVPADGTDRAALAALLTEAADD
FT                       TAPTAVVSLLALDETSGDDAVPAGTTATAALVQALADTGAPAPLWALTRGAVAALPDEQ
FT                       PTAPAQAAVWGLGRIAALELPRHWGGLVDLPADLDERTARRLPAALADAGDEDQLALRA
FT                       TGAYGRRITPAPAPDDAPGTGWQPTGTVLITGGTGALGRHTARWLAAHGAEHLLLLSRS
FT                       GPDAPGAAELTTELTALGARVTLVACDAADREQLTRVLAEVPRDCPLTGVVHTAGVLDD
FT                       GVLTGLTPDRFATVFRAKVASAVLLDELTRDTDLAVFALFSSVAGAVGNPGQAGYAAAN
FT                       AVLDALAARRRAQGLAGTSIAWGAWAGDGMAARHTRPGAEPVGLLDPDLAVPALARAVT
FT                       EPQPTLVLADLQQPRLLESLLALRPSPLLSRLPAARTAARAVQEADRRRAGAAADLRDQ
FT                       LAGTAPADRHAVLLRLVRTTAAAVLGHTGADAIRADKPFRDLGFDSLTAVELSSALAAA
FT                       TGLALPPSLVFDHPSPRALADHLRAELTGDRPESAPAAPPAPVPAADDDPIVVVGMACR
FT                       FPGGVTTPEEFWQLLAEGRDGIDAFPTDRGWDLDVLGRRRPGPQRPPRSAASSYDAAAF
FT                       DPGFFDISPREALAMDPQQRLLLETAWEAVERTGTDPTRLRGSRTGVFVGTNGQDYAGL
FT                       VLRAQEDVEGHAGTGLAASVISGRLAYAFGFEGPAVTVDTACSSSLVALHWAVQALRAG
FT                       ECSLALAGGVTVMTTSTSFAGFTRQGGLAPDGHCKAFSDSADGTGWSEGVGVLVVERRS
FT                       DALRNGHEILAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALANAGLAPGDVDAVEAH
FT                       GTGTVLGDPIEAQALLATYGQDRPADRPLWLGSVKSNIGHTQAAAGAAGLMKMVLALQH
FT                       GTLPRTLHVTEPSTRVDWSAGAVRLLTERTVWPRTDRPRRAGVSSFGISGTNAHVILEQ
FT                       PPAEPTPTAPADRPTRTPAVLPWVVSARSATALDAQLARLRAFAAERPDLPPADVAHSL
FT                       VTSRATFEHRAVLLAAPDGITAAARAEARERSTAFLFSGQGAQRSGMGRELHAAFPVFA
FT                       AAFDEVVAVLDAELATGSGGGVSLREVMWGGGSELLDRTRFTQPALFAVEVALFRLVAS
FT                       WGVGPEFVAGHSVGEIAAAYVAGVFSLVDACRLVVARASLMDALPVGGVMVAVEAAEAE
FT                       VVPLLVDGVAIAAVNGPVSVVVSGVEAAVGQVVDQLVERGRRVRRLAVSHAFHSPLMDP
FT                       MLDAFRAVAEGLEYHQPRIPVVSNVTGEVAAAEELCAADYWVRHVRATVRFADGVRTLA
FT                       ERGATAFLEIGPDGVLSALAAACLFDTDAEVVPALRKGRPEEHTALTAAAQLHVAGVDI
FT                       DWTAVLAGTGGRRIALPTYAFQRERYWPSLAAQAPGDAGGLGLEAGRHPLLGAATTVAG
FT                       SAEILLTGRLSTTAQPWLAVYEADGRTVLPAAVLAELAVRAGDQADCPTVAELTVAAPL
FT                       VLTGAAAQRLQVRVAAPDDTGRRALSVHARPDDSPDSPWTLHATAVLTHDTPQPPAPDT
FT                       GWPPERAVPLDALPTATGPARIAAAWQWGDELCAEIELPEPGPAERAFALHPALLDTAV
FT                       RAGGLLDGDATLDALGWRGLALHAASATALRVRLTPDGTDTWALEATDPQGAPVVSVTG
FT                       LTLGTPTVDRSGAGAADDGATLLDLEWVPAPQAAAPTGGDHLPYAVLGDQLAELDGQLRI
FT                       AGDGPGRVASLAALLDGGAPLPRLVLAPVLGVPTGEGDLPAAVRGTTTAVLELLQRWTA
FT                       DARTADSHLVIVTRGAVAAGAEDVHDLAAAPVWGLVRSAQSEHPGSFLLLDLDPADPAG
FT                       ASRAAAPATLAALLDAGETQAAVRADTLTVARLTRAADGPEATAGHPVRDWDRDGTVLI
FT                       TGGTGGLGGLLARHLVTGHGIKHLLLAGRRGPDAPGARALRDELAALGAEVTVAACDVA
FT                       DRAALDRLLAQLPPEHPLTAVVHTAGVLDDATVGTLTPERLDTVLRAKADAAWHLHDAT
FT                       RDRDLAGFVLYSSVAGVTGGPGQGNYAAGNTFLDALAAHRAAQGLPGLSLAWGPWGQDA
FT                       GMTGTLGAADLARLERSGMPPLTPEQGLALFDAAGARGDGFAVAVRLARGAAAPGADEV
FT                       PAVLRALVRGRRRTAAAAGHAGVLARRLAALDAEQRHQALLDLVRTETAAVLGHSGADA
FT                       VPAERDFNRLGFDSLMAVELRTRLATATGARLPATLVFDHPTPDAVARHLASTLPGGTA
FT                       AGPDRSPLAELDRIAAELSPEGADDATRQGVVGRLRHLLAQWDGTRQDGGGTTVDDRIE
FT                       AASAEEVLAFIDHELGRQADS"
FT      CDS              50747..56947
FT                       /codon_start=1
FT                       /db_xref="SPTREMBL:Q9L4X1"
FT                       /note="polyketide synthase/thioesterase; contains a
FT                       C-terminal thioesterase domain"
FT                       /transl_table=11
FT                       /gene="nysK"
FT                       /function="responsible for the last condensation step in
FT                       the nystatin polyketide backbone synthesis"
FT                       /product="NysK"
FT                       /protein_id="AAF71768.1"
FT                       /translation="MPDEKKLVDYLKWVTKDLHQTRQRLQEVEAGRHEPVAIVGMACRF
```

```
FT                        PGGVRSPEDLWELLSAGRDGIGPFPADRGWDLAALAGDGPGRSATQEGGFLPDAAAFDP
FT                        GFFDISPREALAMDPQQRLLLETAWEAVERSGIDPAGLRGSRTGVFVGTNGQDYAHLVL
FT                        AAQDDMGGYAGNGLAASVLSGRLAFALGLEGPAVTLDTACSSSLVTLHLAAQAVRAGEC
FT                        GLALAGGVTVMTTSSSFAGFSLQGGLAPDGRCKAFAEAADGTGWSEGIGLLLVERLSDA
FT                        QRNGHPVLAVLRGSAVNQDGASNGLSAPNGPSQQRVIRQALAGAGLVPGDVDAVEAHGT
FT                        GTRLGDPIEAGALLATYGQDRPADRPLWLGSVKSNLGHTQAAAGVAGVIKMVLALRHGV
FT                        LPQTLHVDAPSSHVDWESGAVRLLTAPVAWSEGDDRVRRAGVSSFGISGTNAHVILEQA
FT                        PDQPEPTAEETAAAAPGGTAEERAAAPVAPRAVPWPVAARTAGALDAQLVRVRALTTAP
FT                        GRTAADVGHALATARTPFEHRALLVHEGGAVTEVARGAVPTGDRGGLAVLFSGQGSQRP
FT                        GMGRELHARYPVFAAAFDETVALLDARLGTSLRDIVWDQDRTRLDDTRHTQPALFAVEV
FT                        ALYRLLASWGIRPDHVTGHSIGEITAAHVAGVLTLADACTLVAARATAMSELPPGGAMV
FT                        ALEATEDEVRPLLTDDLAIAAVNAPRSVVVAGAEDAALAVRRHFDDLGRRTTRLPVSHA
FT                        FHSPLMDPMLDAFRTALAPLTFAEPEIPVVSNLTGLPATAEELATPHYWVCHVRQAVRF
FT                        GDGVRALADRGVRTFLELGPDGVLSALVRENLPEPGLVAVPVLRKERPEETTVLAALGT
FT                        LWAHGADVDWDAVFAGTRTPQADPVELPTYAFQRARYWPTLGARHGDPADLGQTAAAHP
FT                        LLGAAVTLADADETVLTGRLALPSHPWLGDHRSDGRITVPGVAFAELAVRAGDLSGTPH
FT                        LARLDLPAPLTLGDGDTVTLQVRVGAPDPAGHRPLTVHARLAATEDAPWTTCATGLLAP
FT                        DAPEAPADPIGPADAGWPPRDARPVPVADLDAAATAAGRHYGPHFQGLTGLWRRDGEVF
FT                        AEVALPTATAADRAFGIHPALLATALRATAALDDDHTAGHTPEPTGITGLALHATGATA
FT                        LRVRLTATGPDTVALAAADATGGAVLTADTVTLGSPQDRPAPAPAGHTGQGGLFHLDWV
FT                        PVDPGSRATGTRWAVVGDDELDLGYALHRADETVSAYAASLGGAIGDSGLAPDVFLPV
FT                        VGGPDAGPDAVHAVTARALGLLQEWLNEPRLAGARLVFVTRGAVAVPGETVTDPAGAAV
FT                        WGLLRSAQTENPGSLLLVDLDDAFRSAGMLPHVLTLDEQQLVVRDHAVRAARLARLPEP
FT                        AAGTAPARAWDPDGTVLITGGTGGLGAALARHLVTVRGARHLLLAGRRGPEAPGAGELV
FT                        AELTAQGADVRVAACDVGDRTALDALLATVPAAHPLTAVVHTAGVLDDALIGSLTPDQL
FT                        ATVLRPKADAAWHLHDATRGLDLAGFVLYSSVSGVLGSPGQGNYAAANAYLDALARHRA
FT                        DQGLPALSLAWGPWGRGSGMTASVSDADLERMARGGLPPLTVEDGLALFDAAVGRPEPA
FT                        LVPSRINVAGLRDQQALPALWRDLVPRARRTAATADRSPVTVRERLRHLDETGQEQLLI
FT                        DLVVGYTAGLLGHPDPTAVDPERGFLELGFDSLVSVGLRNQLAEILGLRLPSSIVFDSK
FT                        SPVKLARWLHQELANGPQPGATGPAAADARPAVRSDDTLEGLFYNAVRGGKLVEAMRML
FT                        KAVANTRPMFDTPAELEELSEPVTLADGPGRPRLIFVSAPGATGGVHQYARIAAHFRGS
FT                        RHVSALPLMGFAPGELLPATSEAAARIVAESVLMASEGEPFVMVGHSTGGSLAYLAAGV
FT                        LEDTWDVRPEAVVLLDTASIRYNPGEGNDLDRTTRFYLADIDSPSVTLNSARMSAMAHW
FT                        FMAMTDIQAPAPTAPTLLVRAARALDGFRLDTSSVPADEVRDIDADHLSLAKEHSALTA
FT                        QAIEGWLAELPDPAA"
FT      CDS               57095..58279
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4X0"
FT                        /note="P450 monooxygenase"
FT                        /transl_table=11
FT                        /gene="nysL"
FT                        /function="presumably involved in modification of the
FT                        nystatin macrolactone ring"
FT                        /product="NysL"
FT                        /protein_id="AAF71769.1"
FT                        /translation="MSTPTAPPSLKAEVPPVLRLSPLLRELQSRAPVCKVRTPAGDEGW
FT                        LVTRHTELKQLLHDDRLARAHADPANAPRYVHNPFLDLLVVDDFDLARTLHAEMRSLFT
FT                        PQFSARRVMDLTPRVEALAEGVLAHFVAQGPPADLHNDFSLPFSLSVLCALIGVPAEEQ
FT                        GKLIAALTKLGELDDPARVQEGQDELFGLLSGLARRKRITPEDDVISRLCLKVPSDERI
FT                        GPIASGLLFAGLDSVASHIDLGTVLFIQHPDQLAAALADEKLMRGAVEEILRSAKAGGS
FT                        VLPRYATADVPIGDVTIRAGDLVLLDFTLVNFDRTVFDEPELFDIRRAPNPHLTFGHGM
FT                        WHCIGAPLARVNLRTAYTLLFTRLPGLRLVRPVEELRVLSGQLSAGLTELPVTW"
FT      CDS               complement(58378..58572)
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W9"
FT                        /note="ferredoxin"
FT                        /transl_table=11
FT                        /gene="nysM"
FT                        /function="participates in electron transfer in P450
FT                        monooxygenase systems"
```

```
FT                        /product="NysM"
FT                        /protein_id="AAF71770.1"
FT                        /translation="MRITVDPGRCVGAGQCVLTAPDLFDQDDDGLVTVLAGAADAADPG
FT                        DVRDAAALCPSGAISVAAD"
FT          CDS           complement(58637..59833)
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W8"
FT                        /note="P450 monooxygenase"
FT                        /transl_table=11
FT                        /gene="nysN"
FT                        /function="presumably involved in modification of the
FT                        nystatin macrolactone ring"
FT                        /product="NysN"
FT                        /protein_id="AAF71771.1"
FT                        /translation="MSTEADARTAAPQCPVAFPLRRPGRPFPPPEYATYRGGAGLVRSE
FT                        LPSGPVWLVTRHEDVRAVLTDPRISADPSKPGFPKAGRTGGAPSQYEVPGWFVAMDPPE
FT                        HGRFRKTLIPEFTVRKVRELRPVIQQIVDERIDAMLAAGTSADLVESFALPVPSLVISS
FT                        LLGVPKVDRDFFEDRTRVLVRLSSTDEERDKATQALLRYLGRLIQIKQRRPGDDLISRL
FT                        IAAGTLSRQELSGVAMLLLIAGHETTANNIGLGVVQLLTNPRWIGDDRIVEELLRYYSV
FT                        ADLVAFRVAVEDVEIGGQLIRAGEGIVPLIAAANHDATAFAAPSEFDPERSARSHVAFG
FT                        YGVHQCLGQNLVREEMDIAYRTLFARIPSLTLAVPVEELPLKYDGVLFGLHELPVTWK"
FT          CDS           complement(59830..60888)
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W7"
FT                        /note="putative aminotransferase"
FT                        /transl_table=11
FT                        /gene="nysDII"
FT                        /function="presumably involved in mycosamine biosynthesis"
FT                        /product="NysDII"
FT                        /protein_id="AAF71772.1"
FT                        /translation="MSFTYPVSMPWLQGRELDYVTEAVGGGWISSQGPYVRRFEEAFAA
FT                        YNDVPFGVACSSGTTALTLALRALGVGPGDEVIVPEFTMIASAWAVTYTGATPVFVDCG
FT                        DDLNIDVSRIEEKITPRTKVIMPVHIYGRQCDMDAVLNLAYEYNLRVVEDSAEAHGVRP
FT                        RGDIACFSLFANKIISAGEGGVCLTHDPHLAEQMAHLRAMAFTKDHSFLHKKLAYNFRM
FT                        TNMQAAVALAQTEQLDTILALRRDIEKRYDEALRDIPGITLMPPRDVLWMYDLRAERRD
FT                        ELCAYLAGEGIETRVFFKPMSRQPGYFSADWPALNAARLSADGFYLPTHTGLTAQEQEF
FT                        ITGRIRAFYGVA"
FT          CDS           complement(60909..62429)
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W6"
FT                        /note="putative glycosyltransferase"
FT                        /transl_table=11
FT                        /gene="nysDI"
FT                        /function="presumably responsible for attachment of the
FT                        dGDP-mycosamine to the nystatin macrolactone ring"
FT                        /product="NysDI"
FT                        /protein_id="AAF71773.1"
FT                        /translation="MTLPSGNTRLGWRRRRMHSPGDRAGRVRGARARRPATFRGVLSMG
FT                        ANRRPILFVSYAESGLLNPLLVLAGELSRRDVADLWFATDEKARDEVAAVVDGSPVRFA
FT                        SLGDTVSQMSAVTWDDATYAEVTQRSRFKAHAAVIRHSFAPESRMAKYRRLEEIVEEVE
FT                        PALMVIESMCQFGYELAITKGIPFVLGVPFVPSNVLTSHVPFAKSYTPSGFPVPHSGLP
FT                        AAMSLAQRIENQLFRLRTLGMFLTSDVRKVVEEDNRVRTELGIAPQARQMMARIDHAEQ
FT                        VLCYSVRELDYPFPMHPKLRLVGTMVPPLPQAPDDDGLSDWLSAQKSVVYMGFGTITRL
FT                        TREQVASLVEVARRLDGRGHQVLWKLPRGQQELLPPAAELPDNLRIEGWVPSQLDVLAH
FT                        PNVKAFFTHAGGNGYHEGLYFGKPLVVRPLWVDCDDQAIRGQDFGVSLTLDRPETVDTE
FT                        DVLDKITRVLDQPSFTERAEHFAGLLRDAGGRAAAADLLLGLPALATD"
FT          CDS           62659..66759
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W5"
FT                        /note="polyketide synthase; contains KS domain with
```

```
FT                        Cys->Ser substitution in the active site"
FT                        /transl_table=11
FT                        /gene="nysA"
FT                        /function="presumed loading module of the nystatin PKS
FT                        complex"
FT                        /product="NysA"
FT                        /protein_id="AAF71774.1"
FT                        /translation="MTIGADEDPVVVVGMACRYPGGVAGPEDLWELVRTGRDATTAFPD
FT                        DRGWDLAALAGDGPGRSATREGGFLTGAADFDAAFFGMSPREAVSTDPQQRLVLETAWE
FT                        ALERAGIDPHSLRGSRTGVFVGASGQDYAAVTHASPDDLDGHALTGLAPGVASGRLAYV
FT                        LGLEGPAVTVDTTSSSSLVALHWAVRALRAGECSTALAGGVTVMSTPAAFVGHTRQGGL
FT                        APDGRCKPFSDDADGTAWAEGVGIVVLEHLSTARAAGNPVLAVLRGSAVNQDGASDGLT
FT                        APSGPAQERVIRAALADARLAPADIDLVEAHGTGTRLGDPVEARALLAAYGQDRDPDRP
FT                        LRLGSLKSTLGHAQAAAGIGGVIKTVLTLRHGLMPRIRHLATPTRQVDWSQGAVAPLTD
FT                        HTPWPPADRPRRAGVSSFGISGTNAHVILEEAPPADVPVTRPGTLRPSTVPWPVSAATP
FT                        EALDAQLARLRAHLRTHSDLDPLDVGYSLATGRAALRHRAVLLPPADGTAADAVEHARG
FT                        AAHQRRTAVLFSGQGSQRPGMGRELAARFPVFADALDDALRALDRHLDGPVREVMWGTD
FT                        AALLDRTGWTQPALFAVEVALHRLVASLGVTPDFVGGHSVGEIAAAHVAGVLSLEDACR
FT                        LVAARATLMQALPAGGAMAALEATEDEVAPLLGAHLALAAVNGPTAVVVAGAEDAVRQL
FT                        TARFADRGRRTSRLAVSHAFHSPLMEPMLDAFRDVVSRLTFHQPSIPLVSNLTGELAGS
FT                        EITSAEYWVRHVRDTVRFADGITALAKAGADVLIELGPGGVLSAMARDTLGPDSTTDVV
FT                        PALSKGRPEETAFAGALGRLHTLGVPVDWPAFYAGTGARRVELPTYAFQHVRHWPTPPR
FT                        PNGAGPGALGHPLLGSAVELADGGGTVCSGALSLRTHPWLADHTVAGRVVLPATALLEL
FT                        AVRAGDEAGCDVLHELHLTTPPALPDDAALHVQVHVGPADTTGRRAVTVHTRPDHHPAG
FT                        DWTRCATGTLGSTPPSAAEAATGGTPAAWPPADAEPLDLADHYERLADRGFDYGPTFRG
FT                        LRAAWRRGAEIFADVECPPGTADDAPDHGLHPALLDAARHAAMAVDGTVPVAWHGVRLH
FT                        AVGATALRVRIRPTTTGTLTLTAVDVHGAPVVTVEALTARPLTDEERAAPRTPRQARGE
FT                        TPADARPARPAAARPGPAGEPLPDTTGSHPTAGHLAALPPAARERQLLDLVRTQAAAVL
FT                        GHPGPEAVGTRSVFKELGFDSLAGVELADRLTARTGLRLPATLVFNFPTPERAAHRLGE
FT                        LLAATAPLDPGAYGEELTRFEAIVTNLPQDGPERRAVADRLDAIVSALRQNSPAEVPSS
FT                        DEDIDTVSVDRLLDIIDEEFETT"
FT      CDS               66805..76383
FT                        /codon_start=1
FT                        /db_xref="SPTREMBL:Q9L4W4"
FT                        /note="polyketide synthase"
FT                        /transl_table=11
FT                        /gene="nysB"
FT                        /function="responsible for condensation steps 1 and 2 in
FT                        the nystatin polyketide backbone synthesis"
FT                        /product="NysB"
FT                        /protein_id="AAF71775.1"
FT                        /translation="MQEPQQGQPDQQEKIVDYLRRVTSDLRRARRRIGELESKDNEPIA
FT                        IVGMGCRLPGGVNSPESLWDLVRSGGDAISGFPVDRGWDLETLTGNGDGSSATHEGGFL
FT                        YDAAEFDAAFFGISPREATAMDPQQRLLLEVAWEALERAGIAPTALRGSRSGVFVGSYH
FT                        WGAPSADAATELHGHALTGTAASVLSGRLAYTLGLEGPAVTVDTACSSSLVALHLAAQS
FT                        LRVGESSLAVIGGVTILTEPSVFVEFSAQGGLAPDGRCKAFSDAADGTGWAEGVGVLVA
FT                        ERLSDAQRNGHPVLAVLRGSAVNQDGASNGLTAPNGPSQERVIQQALARTGLTPADIDA
FT                        VEAHGTGTRLGDPIEAQALLATYGQGHTPDQPLWLGSLKSNIGHTQAAAGVAGVIKMVM
FT                        ALRHGHLPPTLHADAPSSHVDWSAGSVRLLTEGQQWPETGRPRRAAVSSFGISGTNAHA
FT                        LLEQAPHPADTADAGDDAAPTEPAGAPAALPWIVSGHSPQALRDQAAALAARVETDPAL
FT                        RPQDIGHTLHTARALLERRAVVVAPDRAELLAATHELAAGRSANAVVEGLADVEGRTVF
FT                        VFPGQGSQWVGMGAQLLDESAVFAERIAECAAALAEFTDWSLVDVLRGVVGAPSLERVD
FT                        VVQPASFAVMVSLAALWGSRGVLPDAVVGHSQGEIAAAVVSGALSLRDGARVVALRSQA
FT                        IGRALAGRGGMMSVALSVDVLEPRLVEFEGRVSVAAVNGPRSVVVAGEPEALDALHARL
FT                        TADDIRARRIAVDYASHSHQVEDLHEELLEVLAELAPRTSEVPFFSTVTGDWLDTARMD
FT                        AGYWFRNLRGRVRFADAVADLLAAEYRAFVEVSSHPVLTMAVLDLIEEAGVTAVATGTL
FT                        RRDQGGAGRFLLSAAEVFVRGVDVDWAGAFEGTGAARVDLPTYAFQRERYWNTRTAADR
FT                        TPADAPMDAEFWAAVEQADVSALTAALGTDEDSVAAILPGLTSWRRARSQRTTLDSWRY
FT                        RVTWTPLAQVPRATLTGTWLLVTTDGIDDTDVAGALESYGAEVRRLVLDEECTDRAVLR
FT                        ERLAGAEDVTGIVSVLAAAEDDAARHPGLTRGLALTVSLVQALGDAEATAPLWFLTRGA
```

```
FT                      FATGPSDPVTRPLQSQIAGVGWTTALEHPQRWGGTVDLPDTLDARAAQRLAAALSGALG
FT                      AEDQLAVRAAGVLARRIVRAGHRAGRPARTWAPRGTTLITGGSGTLAPQLARWLAERGA
FT                      EHVVLVSRRGADAPGAPELIAEAAESGTEVTVAACDITDRDAVAALLADLTADGRTLRT
FT                      VIHAAAAIELSALADTTVAEFADVVHAKVTGARILDELLDDAELDDFVLYSSTAGMWGS
FT                      GVHAAYVAGNAYLSALAEQRRARGLRTTSIHWGKWPDDRARELADPHRIRRSGLEYLDP
FT                      ELALTALQHVLDDDETVIGLMDIDWDTYHDVFTAGRPAHLFDQIPEVRRRLDQASVPDP
FT                      AGPAADGLAARLHGLAAAEQDRLLLTLVRTEAAAVLGHASAESFPERRAFRDLGFDSVT
FT                      AVDLRNRLVAGTGLRLPSTMVFDHPNCAALAAFLKTTALGVPGAAPQQHAATGTPADDD
FT                      PIAVIGMSCRYPGGAATPEELLRLALDGADVISEFPADRGWDARGLYDPDPDRPGHTYS
FT                      VQGGFLHEAAGFDPGFFGISPREAVAMDPQQRLLLETSWEAFERAGIDPASLRGSAAGT
FT                      FFGASYQDYSSTVQNGTGESEAHMVTGTAASVLSGRVSYLLGLEGPAVTVDTACSSSLV
FT                      ALHLACQSLRDGESSLALAGGAAVMATPHAFVGFSRQRALAKDGRCKPFSDTADGMTLA
FT                      EGVGVVLLERLSHARANGHRVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
FT                      LTGADVDAVEAHGTGTKLGDPIEAQALLATYGQDRDAERPLLLGSVKSNIGHTQAAAGV
FT                      AGVIKMVLAMDAGELPGTLHLDAPSSHVDWTAGAVELLRGRTPWPESGRPRRAGVSSFG
FT                      ISGTNAHLILEQAPATEPPADPDRLRDTATDTVVPWPLAAKSPAALRAQAARLLATVEH
FT                      DPDLPPAPVGHALATTRAALEHRAVVVGERREDFLRGLAALSTGASTAGLVSGIAGPDP
FT                      EGAVFVFPGQGSQWWGMGRELLATSEVFRTAIDDCATALAPYVDWSLHDVLAGEGDPAL
FT                      LERVDVVQPALFAMMVGLSALWRSHGVVPAAVVGHSQGEIAAACVAGALSLADAARVVA
FT                      LRSQALPQLSGRGGMMSVSAPVERVTALLAPWQEALSVAAVNGPSSVVVSGDTDALDAL
FT                      HTACQEQGVRARKVSVDYASHGRHVEAVRDELARVLAPVDPRAPEVPFYSTVTGDRVDD
FT                      AAFDGAYWYTNLRQTVRMEEATRALLAAGHRVFIEVSPHPVLAAPIQETQEAVAEATGG
FT                      SAVVLGSLRRDEGGPRRFLTSLAEAHTHGAPVDWTTTFARSAYQPVDLPTYPFQRQDFW
FT                      PEARPATPAAGADASDAAFWQLVENQDLAALADALGVPADDEHTALGTVLPALSAWRAK
FT                      AQARTRIDELRYHVQWTRVAEPAAAPTTGRLLVAVPPDHADAPWVAAALDALGTDTVRF
FT                      EAKGTDRAGWAAQIAQLVEDGEEFTGVVSLLAAAEDLHPDFGSVPLGLGQTLVLVQALG
FT                      DAGLTAPLWCLTRGAVATGRDDALDSPTQGALWGLGRVVALEHPDRWGGLIDLPATLDA
FT                      RAAARLTGLLADPAGEDQLAVRATGVLARRMVHAAPSAPRTGRRWRGRGTCLITGGTGG
FT                      IGGRVARWMAEHGAAHLVLTSRRGPDAPGAAALRAELEALGARVTLAACDVADRDALAA
FT                      LLADLPADQPLTSVFHSAGVADGDARAADLTLDQLDALLRAKLTAAHHLHELTAPLDLD
FT                      AFVLFSSGAAVWGSGGQPGYAAANAYLDALAAHRRSLDLPGASVAWGTWGEVGMATVPE
FT                      VHERLHRQGVRAMEPDHAIGALQQMLEDDDTTLAVTLMDWEAFAPSFTATRPSALFSTV
FT                      PEAVRAVTGDPGTTAGDDVDSATPPLRRHLEELSAAERGRALVEAVRAEASATLGHDTP
FT                      DAIPAGRAFRDVGFDSVTAVELRNRLRTALGLPLPAALVFDHPTPTALAGHLGALLFGT
FT                      APEDAGTGRPDDPDARIREALATVPIGRLRKAGLLDMVLKLADGDATDAPAPEADAPSE
FT                      SLDDMDAEALLRLATENSAN"
FT      CDS             76403..109693
FT                      /codon_start=1
FT                      /db_xref="SPTREMBL:Q9L4W3"
FT                      /note="polyketide synthase"
FT                      /transl_table=11
FT                      /gene="nysC"
FT                      /function="responsible for condensation steps 3 to 8 in the
FT                      nystatin polyketide backbone synthesis"
FT                      /product="NysC"
FT                      /protein_id="AAF71776.1"
FT                      /translation="MSTNPDKYVEALRSSLKEIERLRRQNEQLVAAAVEPVAVVGIGCR
FT                      FPGGVTSPEDLWELVAEGRDVIGPFPQDRGWDLEKLAGGGEGGSLAQVGGFVEDAAGFD
FT                      PGFFGISPREAVAMDPQQRILLEITWEALERAGIDPSTLRGTPTGVFVGTTGQDYGEVI
FT                      KASAEDVEVYSTTGHAASVISGRLSYTLGAEGPAVTVDTGCSSSLVALHWAVQALRGGE
FT                      CSMALAGGASIMATPGPFVAFTAQSGLAADGRCKPFSDRADGTGWGEGAGMLVLMRLSD
FT                      AQREGRPVLAVLRGSAINQDGASNGLTAPNGPSQQRVIRAALDSAHLTAADIDAVEAHG
FT                      TGTTLGDPIEAQALLATYGQDRPRPLWLGSVKSNIGHTQAASGAAGVIKMIMALQRGVL
FT                      PRSLHATEPTTDVDWTAGSVDLLDETVAWPETGRARRAGVSSFGISGTNAHVILEQAPT
FT                      APEEPTTEPTVRPAVVPWALSARTAAALDAQRARLTGHLADTPDADPLDVGYALADGRA
FT                      TFEHRAVLLPDGTELAHGTAGEGPCAVLFSGQGSQRPGMGRELHARFPVFAAAFDEITA
FT                      LLDTHLDRPLREVVWGTDADLLNDTGWAQPALFAVEVALYRLVASLGVTPDFVGGHSIG
FT                      ELAAAHVAGVLSLEDACTLVAARARLMQALPRGGAMLAIRATEDEVTPHLTDDVSIAAV
FT                      NGPTSVVVAGTEEAVAAIGARFTAQDRKTTRLRVSHAFHSPLMDPMLAEFRAVAAGLTY
FT                      HEPRIPVLSNLTGTVAAVADLCSADYWVRHVREAVRFADGVTALTDRGVTTLVELGPDG
```

```
FT          VLSAMAQESLPDGAAAVPLLRKDRPEELSAVTGLARAHVRGVTVRWAGLFDGTGARRAD
FT          LPTYPFQHQRFWPTAARAAQDVTAAGLGAADHPLLGATVELADGAGYLFTSRLSVRTHP
FT          WLADHGVQGRALLPGTAFVELAVRAGDEAGCDRVEELTLAAPLVLPERGGVQLQVRVGA
FT          PDAAGRRTLGIFSRVEDGFDLPWSQHATGVLTAGAGAPDPTFDATVWPPSGAEPVDLTG
FT          AYERLAALGFQYGPAFQGLRAAWRRDTEVYAEVALPDGADTDPAAFGLHPALLDAAQHA
FT          AAYADLGAISRGGLPFAWEGVSLAAAGATTVRARIAPAGEDTVTIAVYDAAGGTVLSVD
FT          SLVSREVPADAPGAAGTVHRDSLFHVEWTPLQGRPGPAPATVAVLGPDPDALADTLRAT
FT          GIRTTAPRDLAALADAEGPVPDLVVTTLTTTPGAPVPDAAHATTAAVLALAQQWLADDR
FT          FADARLVLVTRGATDGTDPAAAAAGGLIRTARTENPGRFALLDLAPDTGRPDPETLATA
FT          LAASHDEPDLAVRGTDVHAARLARVPLATEPTTWNPDGTVLITGGTGGLGAVLARHLVA
FT          THGVRHLLLASRRGPAADGADDLTAELTGLGATVHIAACDVADPAALADLLGTVPAGHP
FT          LTVVVHTAGVVDDGVLGSLTPQRLDTVLRPKADAAWHLHEATRHLDLDAFVLFSSVAAT
FT          LGSPGQANYAAGNAFLDALAARRAATGLPATSLAWGPWTQSVGMTSSLSDLDVERIARS
FT          GMPPLTLEQGTALFDAALAAGPAALAPVRLDLPVLRTQGDIAPLLRGLIRTPVRRTAAQ
FT          VSQTADGLAQRLAGLDAAARREALLELVRTQIAQVLGHADATEVETGRQFQDLGFDSLT
FT          AVELRNALNTATGLRLPATMVFDYPTPHALADHLRDELLGTEAESTTAVPVPTRTAGTD
FT          DPIVIVGMACRYPGGIASPEDLWRLVSQGADATGPFPTNRGWDLDNLYDPDPDRPGRTH
FT          VRAGGFLHDAGSFDADFFGMSPREAMATDSQQRLLLELSWEAVERAGIDPASLRDSGTG
FT          VFAGVMYNDYGTTLTGDEYEAFRGNGSAPSVASGRVSYTLGLEGPAVTVDTACSSSLVA
FT          LHWAAQALRAGECSLALAGGVTVMSTPSTFVEFSRQRGLAPDGRSKAFAEAADGVAWSE
FT          GVGMLVLERQSDAVRNGHEILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGL
FT          STADVDAVEAHGTGTTLGDPIEAQALLATYGRDRDPENPLLLGSIKSNIGHTQAAAGVA
FT          GVIKMVMAMRHGVLPQTLHVDAPSSHVDWSVGAVELLTEQTAWPETGRARRAGVSSFGI
FT          SGTNAHVVIEQSPTAVPATPASADRSVEEPPAVPWALSGKTPDALRDQAARLLAHVEAH
FT          PALRRPVDISYSLIATRTAFDHRAVVLGTDRAEALRALTALAAGETDPAALTGTVRTGRT
FT          AFLFSGQGSQRLGMGRVLYERFPAFAEALDTVLTALDAELGHPLRDIIWGEDAQLVDRT
FT          GYTQPALFAIEVALFRLLEAWGITPDFVAGHSIGEIAAAHVAGVLSLGDACRLVVARAV
FT          LMQSLPEGGAMIAVQATEDEVLPLLTDDVSIAAVNSPTSVVVSGYENATLAVARHFADQ
FT          GRRTTRLRVSHAFHSPLMAPMLDDFRAVVESLTFTAPTTPVVSNLTGELAPAEALCSAD
FT          YWVRHVREAVRFADGIRTLADRGVTTFVELGPDSVLSAMAQESAPEGAGTIPLLRRDRP
FT          EEQAVLAALCHLQVLGVEADWSATFRGLDPVRVDLPTYAFQHRWFWPAARPARPDDVRA
FT          AGLGAAEHPLLGAAVQLPDDDGALFTGRLSLRTHPWLADHTVLGTVLLPGTALVELAVR
FT          AGDETGSGHLEELTLAAPLTLPEDGATLLQVRVGSADDTGRRTVTVHARPDDTADRTWT
FT          LHATGVLATTPPAAAAFDTTVWPPADAEPLTTDDCYAHFTTHRFAYGPAFQGLRAAWRA
FT          GDVLYAEVALPESATDEAAAFGLHPALLDAGLHAALLADDRDTGLPFSWEGVTLHASGA
FT          TALRVRLAPNGPNGLSVTAADPAGNPVATVTRLLARPLDAEQLTIHSALTRDALFHLDW
FT          TPVPLPDTANSAPPALLGPDTAVLADALGDPAVARHATLDDLLAGDTTPPATVLVPLGA
FT          PLDGDTAQHAHALTRSALTLVQQWLATDRLADSRLVFVTHGAVATDDAPPTDLAAAAVW
FT          GLIRSAQTENPGTFTLLDLDTEPDSTTALSRALTLDEPQLLLRAGRARAARLTRTPAPT
FT          TTTHTPWSADGTVLVTGGTGGLGGLVARHLVRSCGVRHLLLLTSRSGVGAAGAAGLVAEL
FT          ESLGARVVVAACDVGDGSAVAELVAGVSESYPLSAVVHAAGVLDDGVVGSLTPERLAAV
FT          LRPKVDGAWNLHEATRGLDLDAFVVFSSVAGVFGGAGQANYAAGNAFLDALMVHRVAGG
FT          LPGVSLAWGAWDQGVGMTAGLTERDVRRAAESGMPLLTVDQGVALFDAALATGSAALVP
FT          VRLDLAALRTRGDIAPLLRGLVRAPLRRTAATGLATGADTGLVQRLGRLDHAQRHEALL
FT          DMVRSSAALVLGHADGNAIDAERAFRDLGFDSLTAVELRNRLRTATGLHLSATMVFDHP
FT          TLSALAEHLRDELFGAVESEVRVPVQALPPTADDPIVVVGMACRFPGGVTSPEDLWRLV
FT          DDGTDAITTFPTNRGWDLDNLYDPDPEHFGTSYTRSGGFLHEAGEFDPAFFGMSPREAL
FT          ATDSQQRLLLESSWEAIERAGIDPLTLRGSATGVFAGVMYSDYGSILGGKEFEGFQGQG
FT          SAGSVASGRVSYALGFEGPAVTVDTACSSSLVALHWAAQALRAGECSLALAGGVTVMST
FT          PSTFVEFSRQRGLAPDGRSKAFAEAADGVGWSEGVGILVLERQSDAVRNGHEILAVIRG
FT          SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSTADVDAVEAHGTGTTLGDPIEAQAL
FT          LATYGRDRDPENPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPQTLHVDAPSSH
FT          VDWSVGAVELLTEQTAWPETGRVRRAGVSSFGISGTNAHVIVEQPALVESPAAEPSGRE
FT          PGVVPLPLSGKSPEALRDQAARLLAGLAERPALRPLDLGYSLATTRSAFDHRAVVLATD
FT          RADAVRALTALAAADADLSAVVGDTRTGRHAVLFSGQGSQRLGMGRELYERFPVFAEAL
FT          DVAIDHLDAALPAQASLREVMWGDDVELLDETGWTQPALFAVEVALFRLVESWGVRPDF
FT          VAGHSIGEIAAAHVVGVFSLEDACRLVAARATLMQALPTGGAMIAIQAAEDEVTQHLTD
FT          DVSIAAVNGPTSVVVSGAESAARTVADRLAENGRKTTRLRVSHAFHSPLMDPMLAEFRA
FT          VAEGLSYATPTLPVVSNLTGRLATADDLCSAEYWARHVREAVRFADGVSTLENEGVTTF
FT          LELGPDGVLSAMAQQSLTGDAATVPALRKDRDEETSALTALAHLHTAGLRVDWAAFFAG
```

FT   SGATRVDLPTYAFQHATYWPTGTLPTAHAAAVGLTAAEHPLLNGSVELAEGEGVLFTGR
FT   LSLQSHPWLADHAVMGQVLLPGTALLELAFRAGDEAGCDRVEELTLAAPLVLPERGAVQ
FT   TQVRVGVADDTGRRTVTVHSRPEHATDVSWTQHATGTLTMGSAPADTGFDATAWPPADA
FT   EPLATDDCYARFTTLGFAYGPVFQGLRAAWRAGDVLYAEVALAESTGDEATAFGLHPAL
FT   LDAALHASLVAHEGEESNGGLPFSWEGATLYATGATALRVRLTPTGTDGRSVAIAVADT
FT   AGRPVAAIDNLVSRRVSGDQLTGAAGLARDALFTLDWNPVPENLVPENPVPENTGGGHA
FT   QDQDGRPAAATVALVGADGTAIAADLTAAGIHTTLHPDLTTLATTDADVPKTVLIPLTG
FT   TGTGTGTESTDGIGTGAAESDASAPSPAEVAHTLSTAALALVQEWTAQERFAGSRLA
FT   FVTTGATAAGGTDVMDVAAAAVWGLVRSAQSEAPDTFVLIDRDPGPAGTHDRTAAAERG
FT   QLLLRALHTDEPQLALRDGGVLAARLARFDTAAALTPPADRAWRLDSTAKGSLNGLALT
FT   PYPAALAPLTGHEVRVEVRAAGLNFRDVLNALGMYPGDDVGSFGSEAAGVVVEVGPEVT
FT   GLAPGDQVMGMITGSFGSLAVDDARRLARLPEDWSWETGASVPLVFLTAYYALKELGGL
FT   RAGEKVLVHAGAGGVGMAAIQIARHVGAEVFATASEGKWDVLRSLGVADDHIASSRTLD
FT   FEAAFAEVAGDRGLDVVLNSLAGDFVDASMRLLGDGGRFLEMGKTDIRAADSVPDGLSY
FT   QSFDLAWVVPETIGTMLAELMDLFRTGALRPLPVRTWDVRHAKDAFRFMSMAKHIGKIV
FT   LTLPRSWKPEGTVLVTGGTGGLGGLVARHLVRSCGVRHLLLTSRSGVGAAGAAGLVAEL
FT   ESLGARVVVAACDVGDGSAVAELVAGVSESYPLSAVVHAAGVLDDGVVGSLTPERLAAV
FT   LRPKVDGAWNLHEATRGLDLDAFVVFSSVAGVFGGAGQANYAAGNAFLDALMVHRVAGG
FT   LPGVSLAWGAWDQGVGMTAGLTERDVRRAAESGMPLLTVDQGVALFDAALATGSAALVP
FT   VRLDLAALRTRGDIAPLLRGLVKAPIRRAAATTPGDTGLAEQLTRLQRAERRDTLLALV
FT   RDQAAMVLGHTSGDGVDPSRAFRDLGFDSLTAVELRNRIGAATGLRLPATAVFDYPTAD
FT   ALAAHLLTELLGPDAESDPDEPGDPTAGPTDDPIVIIGMSCRFPGDIGSPEDLWRLLGD
FT   GADVVTDFPTNRGWDLDNLYDPDPAHAGTSYARTGGFLHDAADFDADFFGMSPREAMAT
FT   DSQQRLLLESSWEAIERAGIDPLTLRDSRTGVFAGVMYSGYGTRLDGAEFEGFQGQGSA
FT   LSVASGRVSYTFGFEGPAMTVDTACSSSLVALHLAAQALRGGECTLALAGGVTVMSIPD
FT   TFIEFSRQRGLAPDGRSKPFSESADGVGWSEGVGMLLLERQSDAVRNGHQILAVVRGSA
FT   VNQDGASNGLTAPNGPSQQRVIRQALASGGLSTADVDAVEAHGTGTTLGDPIEAQALLA
FT   TYGRDRDPENPLLLGSIKSNLGHTQAAAGVAGVIKMVMAMRHGVLPRSLNITEPSSHVD
FT   WSAGAVELLTEQTAWPETGRARRAGISSFGISGTNAHVILEQPEAARHSAPEEADTAEA
FT   AAKAPATAHLPVMPWALSGKTPEALRAQAARLLAHLQQRPELAPADIALSLATQRSQFT
FT   HRAVVLSTDRDEATRALSALATTAASDPSALTGTVTMGRCAVLFSGQGSQRLGMGRELY
FT   ERFPVFAEALDVVIDHLDAALPAQAGLREVMWGDDVELLNETGWTQPALFAIEVALFRL
FT   VESWGVRPDFVAGHSIGEIAAAHVVGVFSLEDACRLVAARATLMQALPAGGAMIAVQAT
FT   EDEVIPHLTDEVAIAAVNGPTSVVISGAEEATQTVAQHFADQGRRTTALRVSHAFHSPL
FT   MMLAEFRAVAEGLSYATPTLPVVSNLTGQVATADELCSAEYWVRHVREAVRFADGVTAL
FT   EAEGVRTFLELGPDGVLAAMARETVADDTVTVPVLRRNMPEERTLLTALGRLHTTGTPI
FT   DWAALLAPTGARPVDLPTYAFQHRPFWPSGPRDTADAAAVGIAGASHPLLNGIVELADE
FT   EGLLFTGRLSLQSHPWLADHAVMGQVLLPGTALLELALRAGDEVGCDHVEELTLAAPLV
FT   LPERGAVQTQVRVGVADTTGRRTVTIHSRPARATTTDSDTHTGTDTPWTQHATGVLVAG
FT   LPATATVPFDATVWPPAHAEPVDLADFYASRAGEGFGYGPAFQGLRAAWRRDGEVFADV
FT   ALPEAGRTEAEAYGLHPALLDAGLHAAWLVAPDGEPTRTGSVPFSWRGVFLAASGASSV
FT   RVRLGRDSDGTLSLAIADTTGAPVASVQALSMRTVSVTALSATAGLARDALFRLDWASA
FT   PEPACQPDDTVTVIPAVAVVGTETSELTSELTAALRAAGADVDVRTTLSTDEPAPALIA
FT   LPLVASDQTGTAEAAPVPAAVHDLTRRALALVQTRLQEQHFADTKFVFVTRGATVGRDV
FT   AAAAVWGLVRSAQSENPGCFALVDLDPDGAVGAAALVAALVSGEPQLAVRGDVLRVARL
FT   VRRPLTEVGAGADGTGDGVGGGSGVSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDL
FT   LLVSRSGERAVGAGELVAELAGVGARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVL
FT   DDGMVGALTGERLSAVLRPKVDAVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAA
FT   ANAFLDALMTRRRAEGLPGLSLAWGPWSLTDGTSGMLADAEADRLTRSGVPPLTAEQGL
FT   ALFDAALATGDATCVPVRLDLSALRAQGEVPPLLRSLIRGRSRRAAAAESATATGLRER
FT   LVGLNPVERQEVLLDLVRGQVALVLGHADADDVHPARAFRELGFDSLTSVELRNRLNTV
FT   TGLRLPATMVFDYPTVEVLVSYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYP
FT   GGVASPDDLWRLVTDGVDAVSPFPTNRGWDVESLYHPDPDHLGTSYTRSGGFLHEAGEF
FT   DPGFFGMSPREALATDSQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAM
FT   LASPEFEGFQGSGSSPSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGEC
FT   GLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDA
FT   VRNGHEILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGT
FT   GTTLGDPIEAQALLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGV
FT   VPRTLHVDAPSSHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNAHVILERPE
FT   AARRPVMETNTVEPSTVPWVLSGKTPEALRAQAAKLLSSIEERPELRLVDVGMSLVTGR

```
FT                     STFEHRAVVLAADRADAARALSAIAADEADAAAATGRVGAGRHAVLFSGQGAQRLGMGR
FT                     ELYERFPVFAEALDVVVDHLDAALPAQAGLREVMWGDDAELLNETGWTQPALFAIEVAL
FT                    .FRLVESWGVRPDFVAGHSIGEIAAAHVAGVFSLEDACRLVAARATLMQALPAGGAMIAV
FT                     QATEDEVTPHLTDDVAIAAINGPNALVVSGVEDAAVEIGARFAAEGRRTTRLHVSHAFH
FT                     SPLMDPMLAEFRVVAEGLSYAAPSLPVVSNLTGQVATADELCSAEYWVRHVREAVRFAD
FT                     GVTALEAEGVRTFLELGPDGVLAAMAGASLTESSLAVPLLRKDRPEEPAALAALAQLHI
FT                     AGARVDWPVLFAGVGAGRVELPTYAFQRGWFWPVGRVGVGGDVGAVGLGSAGHPLLGAA
FT                     VELAAGAGVVLTGRLSLSSHGWLADHAVMGRVFVPGTALLEMVMRAGDEVGCGRVEELT
FT                     LAAPLVLPERGGVRVQVAVDAPDAAGRRGVGVYSCPDGVGQAVWSQHAVGVLASGVADQ
FT                     VGGFGDGGVWPPQGAVSVDAEGCYELFADAGFGYGPVFQGLRAVWRRGEELFAEVALSD
FT                     EVAESADTATGFGLHPALLDASLHASLLSSLEGQSADGGPALPFAWEGVSLFASGATAL
FT                     RVRLAPAGEHAVSVTAVDPTGAPVISIDALRTRRLTLDEVNASHTQLSDALFGVQWTTV
FT                     PSTPAADHPSVAIIGTDHLGLAEALSSSSAGATTTTTAAAYESLDALIAAGPEVSVPDV
FT                     TLIGLTTEDAIAQYVNDHDATVAGQGTIGAGAAAVDAARRLTAEALRTIQAWLADERLA
FT                     ARRLVFVTRGAADGQDVAAAAVQGLVRSAQTENPGTFGLLDLDGTEASTAVLGEALTSD
FT                     EPQLLLRDGHLHAARLTRLASPADTAVPTEWNADGTVLITGGTGGLGAQFARHLVDRYG
FT                     VRNLLLVSRRGPDAPGTTELVAELTAHGAEVAVQACDVADGDAVAALVAGVPDEHPLRA
FT                     VVHTAGVLDDGVIGSLTEERLATVLRPKADAAWHLHEATRGLDLDAFVVFSSVAGVFGG
FT                     AGQANYAAANAFLDALMAQRRAAGLPGLSLAWGPWDQTGGMTGMLSDAEADRLARSGIP
FT                     PLSAEQGLALFDAALALAGTSTPDRAAGSAAASTSGTGDTIAIPAAALVAPVRLDLAAL
FT                     AAQGEVPAILRGLVRTRTRRTAAGGSVTVAGLVNRLSGLTADERRQELLELVRTQAALV
FT                     LGHADPASVDSTAQFRDLGFDSLTAVELRNRLSTATGLRLTATLVFDYPNTDALAEHLR
FT                     DELFGAVESEVRVPVQALPPTADDPIVVVGMACRFPGGVTSPEDLWRLVDAGTDAITTF
FT                     PTNRGWDLESLYDPDPAHLGTSYTRSGGFLHEAGEFDPAFFGMSPREALATDSQQRLLL
FT                     ESSWEAIERAGIDPLTLRGSATGVFAGVMYSDYGSILGGKEFEGFQGQGSAGSVASGRV
FT                     SYTLGFEGPAVTVDTACSSSLVALHLAAQALRAGECTLALAGGVTVMSTPGTFVEFSRQ
FT                     RGLAPDGRSKAFAEAADGVGWSEGVGILVLERQSDAVRNGHEILAVIRGSAVNQDGASN :
FT                     GLTAPNGPSQQRVIRQALASGGLSTADVDAVEAHGTGTTLGDPIEAQALLATYGRDRDP.
FT                     ENPLLLGSIKSNLGHTQAAAGVAGVIKMVMAMRHGVLPQTLHVDAPSSHVDWSVGAVEL
FT                     LTEQTVWPETGRVRRAGVSSFGISGTNAHVILEQPEAVQRLAPGAAETVEPVAIKPSAE
FT                     PSLVPWALSGKSPEALRAQAARLRDFLAERPEPRSIDIGHSLAVTRSQFDHRAIVLVDD
FT                     AKAPADSLAALAALASGVADPAVVSDAVSTGGSAVLFTGQGAQRLGMGRELYGRFPVFA
FT                     EALDVVVDHLDAALPAQAGLREVMWGDDVELLNETGWTQPALFAVEVALFRLVERWGVR
FT                     PDFVAGHSIGEIAAAHVAGVFSLEDACRLVAARATLMQALPTGGAMIAVQATEDEVTPH
FT                     LTDEVAIAAVNGPTSVVISGAEEATQTVAQHFADQGRRTTALRVSHAFHSPLMDPMLAE
FT                     FRAVAEGLSYATPSLPVVSNLTGWLATADELCSAEYWVRHVREAVRFADGITTLEAEGV
FT                     RTFLELGPDGILSALAQQSLAGEAVTVPVLRKDRGEESTALTARAHLHTRGLIEDWQDF
FT                     FAGVGAGRVELPTYAFQRGWFWPVGRVGVGGDVGAVGLGSAGHPLLGAAVELAAGAGVV
FT                     LTGRLSLSSHGWLADHAVMGRVFAPGTALLEMVMRAGDEVGCGRVEELTLAAPLVLPER
FT                     GGVRVQVAVDAPDAAGRRGVGVYSCPDGVGQAVWSQHAVGVLASGAADQVGGFGDGGVW
FT                     PPQGAVSVDAEGCYELFADAGFGYGPVFQGLRAVWRRGEELFAEVALSDEVAESADTAT
FT                     GFGLHPALLDASLHASLLSSLEGQSADGGPALPFAWEGVSLFASGATALRVRLAPAGEH
FT                     AVSVTAVDPTGAPVISIDALRTRRLTLDEVNASHTQLSDALFGVQWTTVPSTPAADHPS
FT                     VAIIGTDPFGLADGLSDALPLVEERGDLAALAASEHPVPDLVLVPVAGTRRTGVPADAE
FT                     GHTDAGTSDMLRSVREATAQVLEQIQQWLADDRFEAARLVFVTRGAVSVGEGGIADLAA
FT                     SAVWGLVRSAQSENPGCFGLLDLDLDLALDSDLAPEVDIERDRDRDPVGGTVQPALAAA
FT                     LHATADEPQLALRGGTVQAARLTRIPAPQTDRAETDPAETDRPEIDTRRPGTVLITGGT
FT                     GGLGGLLARHLVAERGVRSLVLASRSGLAAEGAEKLVADLEALGAVVAVQTCDVADGDA
FT                     VAALVAGVSDEYPLTAVVHTAGVLDDGVIGSLTEERLATVLRPKADAAWHLHEATRDLD
FT                     LDAFVVFSSLAGVLGGAGQANYAAANTFLDALMAQRRAAGLPGVSLAWGPWDRAGGMTG
FT                     TLSDAEADRLARSGVPPISAEQGLALYDAATAGERPLVVPVRLDLAALRGLGDVPALLR
FT                     GLVRTPARRTAAAGAAPSADVLTRQLAGLGGAEQEEVLLRLVRGQAAVVLGHADGSAIG
FT                     AGRQFQELGFDSLTAVEFRNRLNAATGLRLPATLLFDYPTPADVVGHLRGRLGTGEVSG
FT                     AGSVLAALDNLEAVIAGLSLDDAGEHQLVAGRLEVLRAKWADMRSAEGAVDGGADVDIE
FT                     EASDDDMFALLDDELGLN"
FT        CDS          110113..110868
FT                     /codon_start=1
FT                     /db_xref="SPTREMBL:Q9L4W2"
FT                     /note="putative thioesterase"
FT                     /transl_table=11
```

```
FT                          /gene="nysE"
FT                          /product="NysE"
FT                          /protein_id="AAF71777.1"
FT                          /translation="MTTSTEESLWARCFHPAPAAPVRLFCFPHAGGSASFYFPVSAQLS
FT                          SVAEVFAIQYPGRQDRRKEAGVSDLATLADQVYDALRPLLKERPSTFFGHSMGATLAFE
FT                          VARRFEADDGDLVRLFASGRRAPSRVREEAVHRRSDDGIVEELKLLAGTNTALLGDEEI
FT                          LRMILPAIRSDYQAIETYRCPPDVTVRAPLTVLTGDRDPKTSLDEAEAWRGHTTGDFDL
FT                          KVLPGGHFFVSSEAPAIIDLLRAHLAGNG"
FT        CDS              111258..114158
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4W1"
FT                          /note="transcriptional regulator"
FT                          /transl_table=11
FT                          /gene="nysRI"
FT                          /function="transcriptional activator for the nystatin
FT                          biosynthesis genes"
FT                          /product="NysRI"
FT                          /protein_id="AAF71778.1"
FT                          /translation="MRKQSGSSGLLTTLVGRDDELRTLARHAAAARDGRAGLVLLHGPA
FT                          GMGKTSLLRSFTASDVCRGMTVLYGTCGETVAGAGYGGVRELLGGLGLSGGDARRSPLL
FT                          EGLAARALPALTADPAGPDAATGAYPVLHGLYWLAARLMAQRPLVLVLDDVHWCDERSL
FT                          AWIDFLLRRAEDLPLLVVLAWRSEAEPVAPAVLADIAAQRRPTVLGLHPLGPDDIGEMV
FT                          RRVFRTTAAPSFVSRVAAVSGGNPLALARLLDELRAEGVRPDAAGERRAAEVGSHVLAR
FT                          SVRCLLERRPPWVRGVARAIAVLGPECTELLAALAGVPAATVDEALLVLRRAGILAADR
FT                          VDFVHDVVRSAVLDDVAPPTLAELRTNAALLLSDAGRPSEELAGQLMLLPVLDQPWMAA
FT                          VLRDAAAQAESRGAPEAGVRCLYRVLEVEPDNVAVRIQMARALAEINPPEAMRLLKEAL
FT                          SLAGDVRTRAQVAVQYGFTCLAVQESPSGVRMLEDALAELTAELGPEPGPVDRELRTLV
FT                          ESVLLIVGADEKVTIGAVRDRAARLTMPPGDTPAQRQMLAMTTVLTAMDGRDARSAVDQ
FT                          ARRALRAPGVELEPWSLLSASFALSLADEVADAQYALDLMLQYGQDNAAVWTYVLALST
FT                          RALLHHGVGAFPEALADAQTAVEILGEERWADGAVLPRVALATALVDRGEPERAEHVLD
FT                          GITRPRLERFVIEYHWYLQARAYARWVRGDFQGALDLLLACGRSLEESRFSNPAFVPWW
FT                          ADGAVLLATLDRHDQARELAAYGSELAERWGTARGLGLAFMAQGVAAPGRAGIDHLTEA
FT                          VSLLADSPARAMEARAELLLGHAHLKRDDLRAAREHLRAAADLAQRCGAVKLGVDARKL
FT                          LVTAGGRVRRMTASPLDMLTGMERTVADLAVTGASNRAIAEALFVTVRTIETHLTSVYR
FT                          KLGVGGRAELSAVLETRTATSGRQPPAWVSQARGRA"
FT        CDS              114182..117043
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4W0"
FT                          /note="putative transcriptional regulator"
FT                          /transl_table=11
FT                          /gene="nysRII"
FT                          /function="presumable transcriptional activator involved in
FT                          regulation of nystatin biosynthesis"
FT                          /product="NysRII"
FT                          /protein_id="AAF71779.1"
FT                          /translation="MPRSKARNQPTTCTPQCAPDAHGDPTMLLECGREQRLIGDLLHRL
FT                          GQGRPSVLSLTGRPGHAQNALVRWGACRARHDGLRVLRAQATPAERELRYGAVLQLLAV
FT                          LDGPHGSTLDAAIRHDGPPPLPVPGIEEVLRRTGTAPTLVVVEDVQWLDPASLTWLQIL
FT                          LRHLGPDTPLAVLASSCGDTTAFDTDPKAPAVPGPPDTVPVARFVVPALTDRGVAATVR
FT                          AVCGTPGDEEFIAALTSATAGNPAILRDALRAFVDHGLPADADHLPELHALTAGVVGDH
FT                          TVRALDGLPAEVNAVLRALAVCGDLLDFHRVRALAGAHSLSEDRIRTLLASVGLTVSVG
FT                          DKVHIRFPASKARVIEDMPAAERADLYVRAAELTHSCGVNDEDVAHLLLRSSPLGAPWV
FT                          VPLLRRGFAAALRREDHHRACACLSRALQEPLDPRERSLLTLELAAAEAVARPEAGDRR
FT                          LGELVRSTVADTDPTSSGEGVGVRAIDLGFARGNSEWVRRTAGEALPYAGPADREELVA
FT                          LFWLAAVRDDDAPMIPVVPRLPDRPVPPAQAGARAWQLATAGEDADKARKLARIALTGG
FT                          VNESLMMPKLAACAALFATDDNDEAVHGLDTMLTAARSAHLRSMAARIFNLRARIHLCA
FT                          ARLEAAERDLDSAERALPPTSWHPRALPNLIATRILVSMETGRPDRARRLAEAPVPAGG
FT                          EEGVWWPALLLARARVAADDGDWEEALRLSRECGRWLFRRHWANPAMLSWRPLAAEACL
FT                          KLGDVTEARRLRDEELFFADRWGTASARGIARLTTRRLFDDDGDRAVRRIREAAALLRD
FT                          SPARLAYLWSRLSQAGAETAHGDTAAAARSWQAVARMTAAHPASRLATAARTLTVPSVP
```

```
FT                          VATAPPPTAVVPPGWRDLSEAEKDTVLLAARGHGNRQIAEQLAVSRRTVELRLSNAYRKL
FT                          RIGGRKELYLLLEALEGPVADAS"
FT      CDS                 117033..119816
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4V9"
FT                          /note="transcriptional regulator"
FT                          /transl_table=11
FT                          /gene="nysRIII"
FT                          /function="presumable transcriptional activator involved in
FT                          regulation of nystatin biosynthesis"
FT                          /product="NysRIII"
FT                          /protein_id="AAF71780.1"
FT                          /translation="MLLERENELARIRAALDAAEAGDSSLLLINGPLGSGRSALLRRIP
FT                          ELAGDGTRVLRASAAWRERDFPFGIARQLFDHLLSGAGGAGPAERTAGAEHFSRLMDTG
FT                          DRPTGTGPALEVSQAVLQGAQALLADASAERRLLILVDDLQWADGPSLRWLAHLTRRLH
FT                          GLRALLVCTLADGDHRGRYPLVREVAGAAHTVLRLAPLSRDATRVLLAGPQGRPPQDAL
FT                          VRAVYEASRGNPLFLTAFRSALRATGRPPGGDHFGAVRELSPTVLRDRLAGHLRIQPQP
FT                          VREVAVAVAALGDHSDPVLLAQLAGVDEIGFAGARRALVDAGLLARGRDVRFVHGVVRD
FT                          AVDSLLTLDERERSHDDAADLLYRCGRPAEQVAGHLLAVVHPGRPWSEAVLRSAAHNAL
FT                          RAGRPADAARYLRRALLHHRTQDGCRARILVDLATAERALDPDACVRHVSQAVALLDTS
FT                          RDRAAAVLRIPPSLLAAPSPSAVELVRQAAAGLDEPGQRDEEGADELALRLEAWLRHSG
FT                          HENPVELASSVARLRRMGARPPVDSVAERELVAVLLSAGALSGRLSAAEIADTGNRILE
FT                          REPATAAHAHTPLPLVMLSLFVAESVQGVASWLASEQHTRRRYATGADDVLLTAERAFV
FT                          LVTQGRPAAAREHVERALVMDAGDWSEPAVMMFAAVAFELRDPALSERILERIRDRRPA
FT                          GLALTATGQMLQAAVDVHFGRGRDALDTLLACGRRLETVGWRNSALLPWRPYAIGLHQR
FT                          LGETDAALQLAEDELRWAREWGATTNLGRALRLKGWLLQDEGLDLLRESVEILRASSYA
FT                          TELARTLVVLGRRLPGGPEAEAVLREAAGIAAACGVPWLAERAELGLGSAIVPPVATLT
FT                          PSERRVASLVSRGLTNQAIATELGVSSRAVEKHLTSAYRKLGVSGRRELVNALPGR"
FT      CDS                 120268..120900
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4V8"
FT                          /note="putative transcriptional regulator"
FT                          /transl_table=11
FT                          /function="presumable transcriptional response regulator"
FT                          /product="ORF4"
FT                          /protein_id="AAF71781.1"
FT                          /translation="MISAQTAPAGESVGPGLMASLDRDLTIKHANQEFRRRFDDSAGDV
FT                          CGRSFRDLMHPSVQQPLMRQFSRLIEGKRHRFASHVVAVGAQDAAFAGTLTASAVTGKT
FT                          PDIAGILVLMDSSGAADAADAGVVTSQKKFLTEIDARILEGIAAGLSTIPLASRLYLSR
FT                          QGVEYHVTGLLRKLRVPNRAALVSRAYSMGILNVGTWPPKVVDDFIK"
FT      CDS                 121589..122350
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4V7"
FT                          /note="putative repressor"
FT                          /transl_table=11
FT                          /function="presumable transcriptional repressor (DeoR
FT                          type)"
FT                          /product="ORF3"
FT                          /protein_id="AAF71782.1"
FT                          /translation="MDAEGRRRDMLELIRRSGSADVVRLAEEFAVSKETVRRDLNVLEG
FT                          HGLIRRRHGGAYPMVRPGSEAVFVSRTAQPIPEESRIATAAAELLSEAETVFIDEGFTP
FT                          QLIADALPRDRPLTIVTASLPVVSAFATSPQANVLLLGGRVRRGTTATVDHWAVHMLSG
FT                          FVIDLAFLGAEGISRRYGLTTPDPAVAEVKAQAIRVARRPVLAGVHTKFGTASFCRFGE
FT                          VGDLETIVTGAGLPVAEAHRYHLMGPKVLRV"
FT      CDS                 complement(122404..123468)
FT                          /codon_start=1
FT                          /db_xref="SPTREMBL:Q9L4V6"
FT                          /note="putative transcriptional regulator"
FT                          /transl_table=11
FT                          /product="ORF2"
```

```
FT                      /protein_id="AAF71783.1"
FT                      /translation="MAQDSGQTPRSLDHVDQALVHALQITPRASWTRIGSVLGLDAVTV
FT                      ARRWNRLVETGAAWISCHPAPVLAASGQGCLAFVEIDCAPGRLLDVARALAAAPHVVAL
FT                      SHVTGDRDLQLNVMARDPAMLSRWVTHDLAALDGVRAARTHLAGPVHTEGSRWRLRALG
FT                      RHQVARLAADASRHRTDTPAFVLDELDQQLVTALSVDGRATYRALAEQCGAGPDTVRRR
FT                      VQRLFAADMLHARCEVARPLSEWPVTVSFWGQVPAARLREVTRRVTGMREVRLCASVIS
FT                      RHNLHLVAWVRSLDDAQRFEVRLAERAADLTVTERAVALWHMKHGGHLLDEEGYRVGVT
FT                      PLALWREPTDARRG"
XX
SQ       Sequence 123580 BP; 15426 A; 49056 C; 42187 G; 16911 T; 0 other;
         gtatgaccta tttcgccccg tggcgtaagg agtagccggc aggtttcatc cgaaggtggg        60
         gcgcgggagc ggcgagttga ccgtgaggtg cgtgcggtgt tcccggtcgc gccggggggt       120
         gggggccggc aggtggaccg ccgtggtcag cagggcgggg gtgctgtgcc agcggccggg       180
         gaagaccgtg acccggcgac cgcccagcag ggggcccggg accaggagtc gggcggtgaa       240
         ggcgccctgt gtcgggtcga tgacgatctc ggcctcggag aagtccagtt cgcgctgggt       300
         caggggatac cacgccttga agacgctctc cttggcgctg aagagcagcc ggtcccagtg       360
         gacgtccgga cgatgggccg ccagggccac gaggtgcggc cgttccgagg gcagggcgat       420
         ggcgttcagg acgccggccg gcagcggacc gttcggttcg cgtcgatgc tcaccgcggc        480
         cgacagctcc gcggggggaga cggcggcggc gcggtagccg gcgcagtgcg tcatgctgcc       540
         gacgatgccc ggcggccact gcggggcgcc gcgccgattg ggcagtatgg cccggtccgg       600
         gtggccgagc cggcgcaggg cccggcgggc gagatggcgg acggtggtga actcgcgctg       660
         ccgggactcc acggcccggg cgatgacctc gcgttccgag gagagcagcc ggtcgccggg       720
         gcgcggccgg tcgtcgtacg ccgcctcggt ggcgaccgtg gcgggcagga tcagttcgat       780
         caccgcatac ctccggcgaa cggtaagaac aggggggttgc ttcggggcac ggttccgtcc       840
         ttgacggggc gccgtgggcg gccgggtcag ccggccgcgg cgcccgcggt gggagtgtgg       900
         gtgcgggctg cgtgcagccg ggcgtacagg ccctgcgcgc acaggagttg atcgtgggtg       960
         ccctgctcga cgatgcggcc ggcgtccatc acgacgatca ggtccgcgtc gcggatcgtg      1020
         gacaggcggt gcgcgatcac gaagctggtc cggcccgccc gcagggagtt catggcgcgt      1080
         tggatcagga cctcggtccg ggtgtccacg gagctggtgg cctcgtccag cacgaggacg      1140
         gccggtctgg cgaggaaggc ccgggccacg gtcagcagct gcttctcgcc ggcgctgacg      1200
         gtgcccgact cgtcgtccag cacggtgtcg tagccctgcg gcagggtgcg gatgaagcgg      1260
         tcggcacagg tcgcgcgggc cgcctcctcg atgtccgcac ggcaggcacc gggtgcgccg      1320
         tacgcgatgt tctccgcgat ggtgccgccg aacagccagg tgtcctggag caccagcccg      1380
         aagcgggacc gcaggtcgtc gcgggtcatc gtcgcggtgt cggtgccgtc caggaggatg      1440
         cggccggagt ccggttcgta gaagcgcatc aggaggttgc cgagggtggt tttgccggcg      1500
         ccggtggggc cgacgatcgc cacggtgctg cccggttcca cggtcagcga gaggttctcg      1560
         atgaggggcg tgtcggggga gtagcggaag gacacgtcgg tgaactcgac gcggccctcg      1620
         gcgcgggcgg gcgtgccggg ccggagcggg tccggggcct gctcggggggc gtcgagcagg      1680
         gtgaagacgc gctgggcgga ggcgatgccg gactggagcc ggcccgccac cgaggcgatc      1740
         tccacgatcg gctggctgaa ctggcgggcg tagaggatga acgcctgcac gtcaccgagg      1800
         gtcaggtgc cgtttatgac cttccaggcg ccgatgacgg ccaccagcac atagccgagg       1860
         ttggcgacga acatcatgac cggttccatg gcaccggagg cgaactgcgc cttggccgca      1920
         gcccggtaga ccgcgtcgtt gcaggcgtcg aagcgctcct cggccgccgc gcgccggtcg      1980
         aagcccttga tcagcgcatg accggtgcac acctcctcca catgggcgtt gagggtgccg      2040
         ttcgcggacc actgcgcggc gtagtggggc tgcgcgcgct tgctgatccg gccgcgatc       2100
         aacgccgaga ccggcacgct gagcagcatc accacggcca gcgacggcga gatcaccagc      2160
         atcagcacca gcatcgtcaa cagcgagaag atcgaggtga tcagctcggc gagggtctgc      2220
         tggagggtct gttggaggtt gtcgatgtcg ttggtggtgc ggctgagcag ctcaccggcc      2280
         ggctgccggt cgaagtggcg cagcggcagc cgggtcagct tctcccgggc gtcgcggcgc      2340
         agttcgtgga tggtgcgcca caccgcggac gccaccagcc ggccctgcgc cagcatgaac      2400
         agcgacgcca cgacgtagag cgccagcagg accagcagca gccggccgat cgcggcgaag      2460
         tcgatcccgg cgccgggcc cgggacgccg ccgagcacgc cgtcggcgat cagatcggtg       2520
         acccggccga gcagcagcgg gccgaacgcg ttgagcacga tcccgccgac gcccatcgac      2580
         acggcgagtg ccacggagcg cggtgcgga cgcagcaggc cgacgagccg acgtgccggc       2640
         cggggcgcgg tccgctcctc ctcaaggtcg tccggcgagg ccatgggcgg cttcctcctc      2700
         ggtcagctgc gagagcgcga tctcgcggta ggtcgggctg gtgcgcagca gcacgtcgtg      2760
         ggtgccctgg cgaccacgc gcccgcggtc cagcaccacg atccggtccg cgtcgcggcc       2820
         ggcggagatc cgctgcgcca cggtgatcac cgtggcgccc gcggtgtacg gcaccagcgc      2880
         ggtccgcagc gccgcatcgg tggcctggtc gagcgccgag aaacagtcgt cgaagagata      2940
         gatctccggc cggcgcagca acgcccgggc gagggacagg cgttggcgct ggccgccgga      3000
```

```
gacattgccg ccgccctggg tgatctccgc gtcgaggccg tccggcatcc gcgccacgaa      3060
gtcggccgcc tgggcgaccc gcagcgcctc ccacagctcc tcgtcggtgg cgtccggccg      3120
cccgaagcgc agattgctcg ccacggtgcc ggagaacagg tacggccgct gcggcacgaa      3180
cccgacggcg gcggcgagcg tggccgcggt cagctcgcgg acgtcggtgc cgccgacccg      3240
caccgcgccc tcggtggcgt cggccagccg cagcaccaag ttcaacaggg tcgtcttgcc      3300
gctgccggtg ctgccgagca cggcgatccg ctcgcccggc tcgacggtca ggtcgacgtc      3360
ccgcagcacg ggctcctcgg cgcccgggta gcggtacccg gccgcgcaca gttcgatccg      3420
gccggcgggc ccgcgcaccg gctgcggcgc ggccggcggc gccacgctcg acccggtgtc      3480
caggacctcc gcgatccggc cggcacagac ccgggcccgc ggcaccgaca ggaacacgaa      3540
ggcgagcatc acgacggaca tcaggatcag cgagagatag ctcaggaggg cgctgagcga      3600
gccgatcggc atccggcccg cgtcgatccg gtgggagccg gtccacagca gggctacggt      3660
gaaaccgttc atcagcagca gcacgaccgg cagcatcgtc gcgatcagcc gacccacccg      3720
ccgcgacacc acgaggaacg cgtcgttggt ctgcgcgaac cgcgcgcgct cgtggtcgtc      3780
gcggacgaag gaccggacca cccgcacccc ggtgatcgcc tcgcgcagca gccgccccag      3840
ccggtccagg gtcagctgca tccgcgcgta cagggtgccc atccgggcca gcagcaggcc      3900
gaagcagacc gccaccacca gcaccagcgc caccagcagc agtgccagcg gaacgtcctg      3960
gcgcagcgcc agcagcacgc tgcccaggca catcagcggc gcgcagacga cgatgccgaa      4020
gccggtctgg gcgaggttct gcacctgctg cacgtcgttc accgaccggg tcagcaggga      4080
gggggtgccg aaccggccga tctcgcgggc ggagaagtcc aggatgcggc ggaagagcgc      4140
ggaccgcaga tcgcggccca tcgccgtcgc ggtccgggcg gccagcgcgg ccgcaccgag      4200
cgcggcggcg atctgcacca gcgccaccac gcccatcacc acacccagct cggtgatgcg      4260
cccgccgtcg ccgcgcacca ccccctggtc gataagtgcg gcgcccagcg tcggcagcag      4320
caaagtgccc aggatctgga cgagttgaag ggcgacgaga gcggcggtgg cccaggcgta      4380
ggggcgcagc tgtgcccgca gaagtctcaa cagcacggag gaacaccccg gttgacggcg      4440
gcgtgcggcc gcgcggtgga cggagtcggc cggcccgccg acccggtcat tgcaccgcag      4500
tccgctccga aatttcacta gtgttggggg tggcaacgga cttttgacgg cccggtactc      4560
gtgaattccc tagaaagccc gggatgcgtt gacagcatct tccgcggctt gcgagcgtgc      4620
gggtgtctat cgtccggact gcgtattcga cgccaggtcg tggccgagct caagcgtttg      4680
gacggtctct ttcgtgttcg agaagggttt gccatgtcca aacgagcgct gatcaccgga      4740
atcaccggcc aggacggctc ctatctcgcg gagcacctgc tgtcccaggg ctaccaggtg      4800
tggggtctga tccgcggcca ggccaatccc cgcaagtccc gggtcagccg cctcgcctcc      4860
gaactcgact tcatcgacgg ggacctgatg gaccagggca gcctggtctc cgccgtcgac      4920
accgtgcagc ccgacgaggt ctacaacctc ggcgccatct cgttcgtgcc gatgtcctgg      4980
cagcaggccg agctggtcac cgaggtcaac ggcatgggcg tgctgcgcat gctggaagcc      5040
atccgcatgg tcagcggact gtccacctcc cgcacggtca gcccgcgcgg ccagatccgc      5100
ttctaccagg cgtccagctc ggagatgttc ggcaaggccg ccgagacgcc gcagcgcgag      5160
accaccctct tccacccgcg cagcccctac ggcgcggcaa aggcgtacgg gcactacatc      5220
acccgcaact accgcgagtc cttcggcatg tacgcggtct ccggcatgct cttcaaccac      5280
gaatccccgc gccgcggcca ggaattcgtc acccgcaaga tcagcctggc ggtcgcccgc      5340
atcaagcaag gcctccagga caagctggca ctcggcaacc tcgacgcggt gcgcgactgg      5400
ggctatgccg gcgactacgt ccgcgccatg cacctgatgc tccagcagga cgccggcgac      5460
gactacgtca tcggcaccgg gcagatgcac tcggtgcgcg acgcggttcg gatcgcgttc      5520
gaacacgtcg gcctgaactg ggaggactac gtcgtcatcg accccgacct ggtgcggccc      5580
gccgaggtcg aggtgctgtg cgccgacagc gccaaggccc aggaccgcct cggctggaag      5640
ccggacgtcg acttccccac cctcatgcgc atgatggtcg attccgacct ggcgcaggtt      5700
tcccgcgaaa accaatacgg cgacgtgctg ctcgccgcca actggtagca gttctcaagc      5760
tttcgaaaac tagtgaattc ctgccggaat tccgacgaca ctcgccacat ggattccccc      5820
ggtgagtggc gaatccaggt ggcgaatccg aacgtaccgc cgaaacggcg tggagaagtc      5880
ggacgccatt cacgtgcggc cgtcggctgt cgagatgggt tgagttgaga tggacaacga      5940
acaaaaactc cgggattacc tcaagcttgc gacggcgac cttcgacgca cccggcggcg      6000
cgtccacaag ctggagtcgg cggcccagga accggtggcc atcatcggca tgacctgtcg      6060
ctaccccggc ggcgtccgca gccccgaaga cctctggcgc atggtcgagg ccggcgagca      6120
cggcgtcacc ccgttcccca ccgaccgcgg ttgggacctg gaggcgctgg ccgccgcgcc      6180
gaccgcctcc ggcggattcc tgcacgacgc acccgacttc gacgcggact cttcggcat      6240
ctcgccgcgc gaggcggtcg ccatggaccc gcaacagcgc gtcgtcctgg aatccgcctg      6300
ggaggcgttc gaacgcgccg gcatcgaccc gacgtccgtg aagggcagcc gcaccggagt      6360
cttcatcggc gcgatggccc aggactaccg ggtcggcccc gccgacggcg ccgagggctt      6420
ccaactcacc ggcaacaccg gcagcgtgct gtccggccgc atctcctaca ccttcggcac      6480
ggtcggcccc gccgtcaccg tcgacaccgc ctgctcctcc tccctcgtcg ccgtccacct      6540
cgccacccag gcgctgcggg ccggcgagtg caccctcgcc ctcgccggcg gcgtcaccat      6600
```

```
catgtccggc cccggcacct tcatcgaaat gggccgccag ggcgggctct ccgccgacgg   6660
ccgctgccgc tccttcggcg acaccgccga cggcaccggc tgggccgaag gcgtcggcat   6720
cctcgtcctg gaacggctgt ccgacgccgt ccgcaacggc cacgagatcc tcgccgtcgt   6780
ccgcggcacc gccgtcaacc aggacggcgc ctccaacggc ctgaccgccc ccaacggccc   6840
ctcccagcag caggtcatcc agcaggccct ggtcaacgcc cgactcgccg ccggggacat   6900
cgacgtcgtc gaggcgcacg gcaccggcac caccctcggc gaccccgtcg aggcccaggc   6960
cctgctcgcc acctacgggc agaaccgccc ggcggaccgg ccgctgctgc tgggctcggt   7020
caagtccaac ctcagccaca cccaggccgc cgccggcgtc gccggcgtga tcaagatggt   7080
catggcgatg cggcacggca ccctgccgcg caccctgcac gccgaggagc ccacccacca   7140
cgtcgactgg tcgcagggcg ccgtgcggct gctgaccgac accaccgact ggcccgccac   7200
cggggcgccg cgccgcgccg ccgtctcctc cttcggcatc agcggcacca acgcccacac   7260
catcatcgag caggcccccg aaccgcagcc cgaggacgcc gcgaccgcgc aggacgacgc   7320
cgccggcagc acgccggcca ccgcccccgt agtgcccggc gtcgtaccgg tcctgctctc   7380
cggccgcacc ccggacgccc tgcgcggcca ggccgcggcc ctgcgcgccg ccctcgacac   7440
cggccggcgg cccgacctgc tcgacctcgc acactccctc gccaccaccc gcgccgggtt   7500
cgagcaccgc gccgtcctcc tcgccaccga ccaccccgcc ctgaccgacg gcctcaccgc   7560
cctcgccgac gccgacgacc cggccgccgc ccccgcctgg atcaccggca ccacccgggc   7620
cgagacccgg ctcgccgtcc tgttcaccgg ccagggcgcc caacgcctcg gcgcgggacg   7680
ggaactcgcc gcccgtttcc cggcgttcgc caccgccctc gacgcggcgc tcgacgcctt   7740
cacccccgcac ctcgaccgcc ccctgcgcga ggtcctgtgg ggcaccgacg ccgccctgct   7800
cgaccgcacc gcatacgccc agccggccct cttcgccgtc gaagtggcgc tctaccggct   7860
gatcgaatcg ttcggcgtcc gccccgacca cctcgccggc cactccgtcg gcgagatcgt   7920
cgccgcgcac ctcgccgggg tcctctccct ggccgacgcc gccaccctcg tcgccgcccg   7980
cggtcgcctg atgcaggcgc tgcccgacgg cggggcgatg atcgccgtcc aggcgtcgga   8040
agccgacgtc gccccgctgc tcgccgggca cgaggaccag gtcgcgatcg ccgccgtcaa   8100
cggcccctcc gccgtcgtcc tgtccggcgc cgaagccacc gtcaccgcgc tcgccgaaca   8160
gctcgccgcc gacggccgca agacccgccg gctgcgcgtc tcgcacgcct tccactcgcc   8220
gctcatggag ccgatgctcg acgccttccg cgccgtcgtc gaagacctca cgctccagcc   8280
gccgctcctg ccggtcgtct ccaacctgac cggcaagccc gccaccgtcg cccaactcac   8340
ctccgccgac tactgggtcg accacgtccg gcacgccgtc cgcttcgccg acggcatcga   8400
ctggctcgcc cggcacgaca ccaccgcctt cctcgaactc ggccccgacg gcgtgctgtc   8460
cgccatggcc caggactgcc tggacgccgc cgacgcagac gccgtcaccc tccccgccct   8520
gcgcgccggg cgccccgagg agcacaccct caccaccgcc ctcgccggtc tgcacgtcca   8580
cggcgccacc ctggactgga ccggctgctt cgccggcacc ggcgcccgcc gcaccgacct   8640
gccgacctac gccttccagc gccgccgcta ctgcccaag gccctccaga gcggcaccgc   8700
cgacctgcgc tcggtcggcc tcggtgccgc ccaccacccg ctgctctccg ccgccgtctc   8760
cctcgccgac gcaggcggca ccctgctcac cggccgcctc tcccggcaga cccacccctg   8820
gctcgccgac cacaccgtcc gcggcaccac cctgctgccc ggtaccgcct tcctcgaact   8880
cgccgtccgc gccggcgacg aggtcggctg cgaccgcgtc gaggaactca ccctcgccgc   8940
accgctcctg ctgcccgaac agggcggcgt ccaggtccag ttgtggatcg gcaaccccga   9000
cgtgtccggt cgccgcaccg tcaacgtcca cgcccgcccc gacaccggcg acgacacccc   9060
ctggaccgcc cacgccaccg gcgtcctcac caccgccgac gcctccgcc agctcccggc   9120
ttcgtccgag cagggcggca cccccctcgc cggcgacccc caccccgccc tcgacgcggc   9180
ccagtggccc ccggccggcg ccgaaccgct gccgctggac ggccactacg accgcctcgc   9240
cgacggcggc ttcggctacg gcccggtctt ccagggcctg cgcgccgcct ggcgcggcgg   9300
cgacgtcgtc tacgccgagg tcgagctgcc cgaggccggc cggtccgacg ccgaggcgtt   9360
cggcctccac cccgccctgc tcgacgccgc cctgcacgcc gcgcccttca ccggcctcgg   9420
cgaacgcggc cggggcggcc tgccgttctc ctgggagggc gtctccctcc acgccggcgg   9480
cgccaccacc ctccgcgtcc gcctgacccc ggtcgccgac gacgcgctcg ccctgaccgt   9540
cgccgacggc accggcgcgc ccgtgctgtc cgtcgactcg ctcgtcctgc gcagcgtggc   9600
gacccaacag ctcgacacgg ccgccgccgt cgcccgtgac gccctcttcc gcctcgactg   9660
gaccccgtc cagccgaccg ccaccgaccc cgggcccgtc gccctcctcg gcgccgaccc   9720
cttcggcctg ctcacccacg ccggattcgc cgacgccccg gcatacccgg acctcgccgc   9780
cctcgccgcg gcggacggcc cggtcccgac caccgtcgtg ctgtccctcg ccggcaccgg   9840
ggacgacgcg gccgacccgg cccggtccgc acaccgctgc gccgcggagg ccctcgccgc   9900
cgtacagacc tggctcgacc accatgagcg cttcgccgcc gcccgcctgg tcttcgtgac   9960
ccgtggtgcg acggtcgggc gtgatgttgc tgctgctgcg gtgtggggtc tggtgcgttc  10020
ggcgcagtcg gagaatccgg ggtgttttgc tctggtcgat ctggatccgg atggtgcggt  10080
gggtgcggct gcgctcgtcg ctgcgttggt cagtggtgag ccgcagcttg cggtgcgcgg  10140
tgatgtgttg cgggtcgcgc gtctggtgcg gcggccgctc accgaggtcg gtgcgggtgc  10200
```

```
tgatggcacc ggggatggcg tcggggatgg ctctggtgtg tcgttctcgg gtgagggtgc    10260
ggtcctggtc actggtggta cgggtggtct gggtgcggtg ttggcgcgtc atctggtggc    10320
cgagtatggg gtgcgggatc tgctgttggt cagtcgcagt ggtgaacgtg ccgtgggtgc    10380
tggggagttg gtggcggagc ttgcgggtgt gggtgcgcgg gtgcgggtgg ttgcgtgtga    10440
tgtgaccgat cgtgccgcgg tggtggagtt ggttggcggg catgcggtgt ccgcggtggt    10500
tcatgcggct ggtgtgctgg atgacggcat ggtgggtgcg ttgaccgggg agcggttgtc    10560
cgcggtgctg cggccgaagg tggatgctgt ctggcatcta catgaggcga cccgcggcct    10620
ggacctggac gcgttcgtcg tcttctcctc cctcgccggg gtcttcggca gtcccggcca    10680
ggccaactac gcggccgcga acgccttcct ggacgcgctg atgacgcggc gccgggcgga    10740
gggactgccc ggcctgtcac tcgcatgggg accgtgggag cagtcgggcg gaatgacggg    10800
caccctgacg gacgtcgacg ccgaacggct ggcccgctcc ggtgtcccgc cgctctccgt    10860
ggcgcagggc ctggccctct tcgacgctgc cgtggccggg accgacgcca cctgcgttcc    10920
ggtccgcctg gacctccccg tcctccgcgc acggggtgaa gtgccgccgc tgctgaggtc    10980
gttgatccgc gtccgggcgc gccgagccgc cgtcgccggc tccgccaccg cgggcaacct    11040
cgcccagcgc ctgcgccgcc tggacgagga cggccgcgac gagatggtcc tggacctcgt    11100
ccgcggtcag gtcgccctcg tcctcggcca cgcgaccggt ggcgacgtcg acgccggccg    11160
tgccttccgc gacctcggct tcgactcgct gaccgccgtc gaactgcgca accgcctcaa    11220
caccgtcacc ggcctgcgcc tgcccgccac cctggtcttc gactacccga ccgtccggca    11280
cctcgccacg tacgtcctgg acgagttgtt gggcacggat gccgaggtgg cgaccgtgca    11340
gccggccgcg gttgcggtgg cggacgatcc gatcgtcatc gtgggcatgg cctgccgcta    11400
ccccggtggc gtcagctccc ccgaggacct gtggcgcgtg ctcaccgaag gcaccgacgc    11460
cgtctcgggc ttcccgacca accgtggttg ggacgtcgaa tccctctatc acccggaccc    11520
tgaccacccg ggtacctcct acacgcgctc gggtgggttc ctgcatgagg cgggggagtt    11580
cgatccgggg ttcttcggga tgagtccgcg ggaggcgttg gcgaccgatt cccagcagcg    11640
gttgttgttg gagtcgtcgt gggaggcgat cgagcgggcc ggtattgatc cggtgagttt    11700
gcggggtagt cggacgggtg tgttcgcggg ggtgatgtac agcgattaca gcgcgatgtt    11760
ggcgagtccg gagttcgagg gtttccaggg cagtgggagt tcgccgagtt tggcgtcggg    11820
tcgggtggcc tacacgttgg ggttggaagg cccggcggtg acggtggata cggcgtgttc    11880
gtcgtcgttg gtggcgatgc actgggcgat gcaggcgttg cgtagtggtg agtgtgggtt    11940
ggcgttggcc ggtggtgtga cggtgatgtc gacgcctgcg gtgtttgtgg actttgctcg    12000
gcagcggggt ttgtcgccgg atggccggtg caaggcgttt gcggatgcgg ccgatggtgt    12060
gggctggtcc gagggcgtcg gcgtgttggt cctggagcgg cagtcggacg cggtgcgcaa    12120
tggtcacgag attttggctg tggtgcgggg ttcggcggtc aaccaggatg gtgcgtccaa    12180
tggtttgacg gcgcccaatg gtccgtcgca gcagcgggtg atccggcagg cgttggccag    12240
tggtggtctg acggccggtg acgtggatgt ggtggaggcg catggtacgg gtacgacgct    12300
cggtgatccg atcgaggcgc aggcgttgtt ggcgacgtat gggcgggatc gtgagcctga    12360
gcggccgttg ttgttgggtt cggtgaagtc gaatctgggg catacgcagg ctgctgcggg    12420
tgtggcgggc gttatcaaga tggtgttggc gatgcggcat ggtgtggtgc cgcggacgtt    12480
gcatgtggat gcgccttctt cgcatgtgga ctggtccgag ggtgcggtgg agctgctcag    12540
tgagcaggcg gcctggccgg agacgggtcg ggtgcggcgg gcgggtgtct cctccttcgg    12600
catcagcggt accaatgtgc atgtcatcgt cgagcaggcg ccgggcgcca aggcgatcgc    12660
cgcggccggt gcggcgcggc gcacgccggg tgccgtgccg gtgctgctct cggggcgtgg    12720
ccggagtgcc ctgcggggcc aggccgcccg cctgctcgga cacctccagg cccgacccga    12780
cgccgaactc gtcgatgtcg cactgtcgtt ggcgaccacc cggtcccgct tcgagcagcg    12840
ggccgccgtc gtggcgcagg accgcgacca gctgatcgcc tcgctggggg cgctggccgc    12900
cgaccgcccc gaccccgccg tcgtcgaggg cgaggccgcc ggacgcggcc ggaccgcggt    12960
gctgttcacc ggacagggca gccagcgggc cgccatgggg cgtgaactcc acgaggtgca    13020
gccggagttc gccgcggcgt tcgacgcggt gtgtgccgtt ttcgacccgc tgttggaccg    13080
gccgctgcgc gaggtggtgt tcgccgagga cggcagcgac gaggccgcac tgctggacga    13140
gaccggttgg acgcagccgg ctctgttcgc cgtcgaggtg gcgctgttcc ggctggtgga    13200
gagttggggt gtccgtccgg acttcgtggc cggccattcc atcggtgaga tcgcggcggc    13260
gcacgtcgcc ggggtgctga cgttggagga cgcctgccgt ctggtggccg cgcgggcgac    13320
gctgatgcag cgcctgccga ccggcggcgc gatgatcgcg atccaggcca ccgaggacga    13380
gatcgcggcg cacctcgacg acacggtggc gatcgccgcc gtcaacgggc cgcagtccgt    13440
ggtgatctcc ggtgacgagg aggccgccga aacgatcgcc gccacgttcg ccgaacgcgg    13500
gcgcaagacc aagcggctgc gggtgagcca tgccttccac tcgccgcgga tggacgggat    13560
gctggacgcc ttccggatcg tcgccgaggg gctgacctac cgggcgccgc gcatcccgct    13620
cgtctccgac ctcaccggcc ggcgcgccga cgatgcggag gtgtgcaccg cggagtactg    13680
ggtgcggcac gtccgagagg ccgtgcggtt cgccgactgc gtgcggacgc tgcgcgacgc    13740
cggggccacc accttcctgg aactgggctc cgacggcctg ctgaccgcga tggccgagga    13800
```

```
caccctcggt gacgaccacg acgccgaact ggtgccgatg ctgcgcgccg ggcgcgccga    13860
ggaactggcc gcggccaccg ccctggcccg cctccaggtg cgcggcgtgg acgtggactg    13920
ggcggcgtac ctcgccggca ccggcgcccg acgcaccgac ctgccgacct acgccttcca    13980
gcacgcgtac tactggccgc agctgccgac cccggccgcc gccctcgccg ccgccgatcc    14040
cgccgaccag cagctgtggg ccgctgtgga gcgcggcgac gcccgcgaac tcgccgacat    14100
cctcggcctg ggcgaacagg acctcacgcc gctggactcc ctgctgcccg ccctcacctc    14160
gtggcggcgc ggcaaccagg agaagcacct cctggacacc ctgcgctacc gcgtggagtg    14220
gacacgactg agcaagccga ccgccccggt cctcgacggc acctggctgc tggtcgcctc    14280
cgacgccacc gcggccgacc agccagccct cctcgacggc ctggccgacg ccctcggctc    14340
gcacggcgcg cgggtgcgtc gcctgcttct ggacgactcc tgcgcggacc gcgcggtgct    14400
cgccgaacga ctggcgcgga ccgccgacgt ggacgccgcg acccaggtgc tgtccgtgct    14460
gccgctcgac gagcgggacg ccgacgactg cccgccgctc acccgcggac tggcgctgac    14520
cgtcgcgctc gtccaggccc tcgccgacac cggcgcccag ggccggctgt ggaccgccac    14580
ccgcggcgcc gtctccacca accccgccga cccggtcacc caccccgtcc aggccgctgc    14640
ctggggcctg ggccggggcg tcgccctgga gcacccacgg ctgtggggcg gcctggtcga    14700
cctgccgcag gtcttcgacg agcgggccgg acagcggctc gccgggatcc tcgccgtcaa    14760
ggacgcaccg gacggcgagg accaggtggc gctgcgggcc accggagtct ccggccgccg    14820
gctcgtccgc cacaccgtcg aagcgctgcc cacggccgcg gagttcaccg ccaccggcac    14880
tgtcctgatc actggtggca ccggtggcct gggcgccgag gtcgcccggt ggctggcccg    14940
cgccggcgcc cagcacctcg tcctgaccag ccgccgcggc ccggacgcgc cgggcgccgc    15000
cgaactccgg gccgaactgg agggctacgg gccgtcggtg tccgtcgtcg cctgcgacgt    15060
cgccgaccgg gacgcgctcg ccgccgtcct caccgcactg cccgaggaac tgccgctgac    15120
cggtgtcgtg cacaccgcag gcgtcggcca ctacggcccg ctggacaccc tgagcaccgc    15180
cgagttcgcc ggcctcaccg ccgccaagct cgccggcgcc gcccacctcg acgccctgct    15240
cgccgaccgc gaactggact tcttcgtcct cttcggctcc atcgccggtg tctggggcag    15300
tggcaaccag agcgcctacg gcgccgccaa cgcctacctc gacgcgctcg ccctgcaccg    15360
ccgcgcccgc ggcctcgccg cgacctccgt cgcctggggc ccgtgggccg aggccggcat    15420
ggccgccgac gatgccgttt ccgagaccct gcgccgccag ggcctcggcc tgctcgaccc    15480
ggccccggcc atgaccgagc tgcgccgcgc cgtcgtccgg caggacgtca ccgtcaccgt    15540
cgccgacgtg gactggcagc gctacgcacc gctgttcacc tccgcccggc ccagcgccct    15600
gatcgccggc ctgcccgagg tccgcgccct cgccgccgac gagcgcaccg agcaggacgc    15660
caccggcgcc tccgaggtcg tcacccgcgt ccgcgccctg gccgaacccg agcaactgcg    15720
cctgctgacc gacctcgtcc gcaccgagtc cgccaccgtc ctcggccaca gctccgccga    15780
cgccgtgccc gagggccgcg ccttccgcga cgtcggcttc gactcgctga ccgcggtcga    15840
gctccgcaag cgcctgggcg ccgcgaccgg gctgtccctg cccagcacca tggtcttcga    15900
ctacccgaca ccgctggaac tcgcccagta cctgcgggcg gagatcctcg gcgcggtgct    15960
ggaagtcgcc ggcccggtcg ccaccggcgg cgccgacgac gagccgatcg ccatcatcgg    16020
catggcctgc cgcttccccg gcggcgtcag ctccccggaa cagctgtggg acctggtcgc    16080
ctccggcacc gacgcgatca gcgagttccc cgtcaaccgc ggctggcaga ccgggcacct    16140
cttcgacccg gaccccgacc ggcccggcac cacctactcc acccagggcg gcttcctcca    16200
cgaggccgac gagttcgacc ccaccttctt cggcatctcg ccccgcgagg cgctggtcat    16260
ggacccgcag cagcggctcc tgctggagac cacctgggag tccttcgagc gcgccgggat    16320
ccgcccggaa accctccgat ccaccctgac cggcaccttc gtcggctcca gctaccagga    16380
gtacggcctg ggcgcgggcg acggcaccga gggccacatg gtcaccggca gcagccccag    16440
tgtgctctcc ggccgactgt cgtacgtctt cggtctggaa ggcccggcgg tcaccgtcga    16500
caccgcctgc tcgtcctcgc tcgtggcgct gcacctggcc tgccagtcgc tgcgcaacgg    16560
cgagagcaac ctggccgtcg ccggcggcgc cacgatcatg acgacgccca acccgttcat    16620
cgcgttcagc cggcagcgcg ccctcgccaa ggacggccgc tgcaaggcgt tctccgacga    16680
cgcggacggc atgacgctcg ccgagggcgt cggcgtcgtc ctcgtcgagc ggctctccga    16740
cgcgcagcgc aacggccacc cggtcctggc cgtcctccgc ggctccgcca tcaaccagga    16800
cggcgcctcc aacggcctga ccgcgcccaa cggcccgtcc cagcagaggg tcatccgcca    16860
ggccctcgcc aacgcccgcc tcgcgcccgg ggacatcgac gccctggagg cgcacggcac    16920
cggcacaccg ctcggcgacc ccatcgaggc ccaggcactg ttcgccacct acggccgcga    16980
ccgcgacccc gagagcgcgc tgctgctcgg ctcggtgaag tccaacatcg gccacaccca    17040
gtccgccgcg ggcatcgcca gcgtcatcaa gatggtcatg gcgctgcgcc actccgaact    17100
gccgccgacc ctgcacgccg acgcgccgtc ctcgcacgtg gactggtcgg ccgggacggt    17160
ccggctgctg acccaggcgc gcgcctggcc ggagaccggt cgcccgcgcc gggccgcggt    17220
gtcctcgttc ggcatcagcg gtaccaacgc ccatgtcctg ctggagcagg cgcccgtcgc    17280
ggacaccccg gccgaggagc ggcccgccgt ggcgccggtc ccgatcgccg ccggcgtcgt    17340
cccgtgggtg gtcaccgccc gcagcgccgc cgccctgcgc ggccaggccg agcgcctcct    17400
```

```
cgcgcacgcc gaaaccgtcg gaaccgccct gccggccgcc ggaccgctcg acatcggcct   17460
gtcgctggtc tccgcgcgcg cccgtttcga gcaccgtgcc gtcgtcgtcc cgcccgcggg   17520
caccgacccg ctggccgcgc tgcgcgccgt cgcgacggac gggccctcgc ccgtggtcgc   17580
ccgtggcgtg gcggacgtcg agggtcggac ggtgttcgtg ttccccggtc agggttcgca   17640
gtgggtgggg atggggtccc aactccttga tgagtcggcg gtgttcgcgg agcggattgc   17700
cgagtgtgcg gcggcactcg ccgagttcac cgactggtcg ctggtcgatg tgctgcgggg   17760
tgtggtgggt gcgccgtcgt tggagcgggt cgatgtggtg cagccggcgt cgttcgcggt   17820
gatggtgtcg ttggctgcgt tgtggcgttc ccgtggtgtg ttgccggatg cggtggtggg   17880
gcattcgcag ggtgagatcg ctgctgcggt ggtgtcgggt gcgctgtcgt tgcgggacgg   17940
ggcgcgggtg gtggcgctgc ggagtcaggc cattggtcgt gcgttggcgg ggcggggagg   18000
gatgatgtcc gtcgcgctgt cggtggacgt gctcgaaccg cggttggtcg agttcgaggg   18060
gcgggtgtcg gtggccgccg tcaacggccc gcgctccgtc gtggtcgccg gcgagcccga   18120
ggcgctggac gcgctgcacg cccggctgac cgccgacgac atccgggccc gccggatcgc   18180
ggtggactac gcctcgcact cgcaccaggt cgaggacctg cacgaggaac tgctggaggt   18240
gctggcggag ctggcgccgc gcacgtcgga ggtgccgttc ttctcgaccg tgaccggcga   18300
ctggctggac accgcgcgga tggacgccgg ctactggttc cgcaacctgc gcggacgggt   18360
gcggttcgcg gacgcggtgg cggacctgct ggcggcggag taccgcgcat tcgtcgaggt   18420
cagctcgcac ccggtgctgt cgatggcggt gcaggaggcg atcgacgagg ccggcgtgcc   18480
ggccgtcgcc gccggcaccc tgcgccgcga ccagggcggc accgaccgct tcctgctgtc   18540
ggccgccgag gtcttcgtgc gcggtgtgga cgtggactgg gcggggctgt tcgaggggac   18600
cggtgcgtcc cggatcgacc tgcccaccta cgccttccaa cacgaacacc tgtgggccgt   18660
cccgcccgcc ccggaggccg tcgccgccgc cgacccggac gacgcggcct tctggaccgc   18720
ggtcgaggac ggtgacgtct ccgcgctgac cgccgcgctc ggcaccgacg aggactccgt   18780
cgccgccgtg ctgcccgccc tgacctcctg gcgccgggcc cgccgcgacc gctccaccgt   18840
ggacgcctgg cgctaccgcg tcgcctggaa acccctcggc ggcaccctgc cgcacccgtc   18900
cctgaccggc acctggctgc tggtcaccgc cgacggcatc gacgacaccg atgtggcagg   18960
ggcgttggag acctacggcg ccgaggtgcg ccggctggtc ctggacgagg agtgcgtcga   19020
ccgcgccgtc ctgcgggagc ggctggccgg cgcggaggac gtgaccggca tcgtctccgt   19080
cctcgccgcc gccgagcgga ccgacgcggt accgggcacc tccctggtgc tcggcaccgc   19140
cctgaccgtg gcactgatcc aggccctcgg cgacgccgaa atcgacgctc ccgtatgggc   19200
gttgacccgc ggcgcggtct ccaccggccg ggccgacgag ctgaccgcgc ccgtccaggc   19260
acaggtcacc ggcatcggct ggaccgcggc gctggagcac ccgcagcgct ggggcggcac   19320
cctcgacctg cccgccgccc tcgacgcccg ggccgcccag cggctcgccg ccgtgctgtc   19380
cggcgccctc ggcagcgacg accagctggc catccggccc tccggggtct tcacccgccg   19440
catcgtgcgg gccgaggcca ccgccgggcg gcccgccggc acctggacgc cgcgcggcac   19500
cacactggtc accggcggct ccggcaccct cgccccgcac ctcgcccgct ggctggccca   19560
acgcggcgcc gagcacctgg tcctgatcag ccggcgcggc acggccgccc cgggcgccgc   19620
cgaactcgtc gcggaactgg ccgagtcggg caccgaggcg accgtcgccg cctgcgacat   19680
caccgaccgc gacgcggtcg ccgcgctgct ggccgacctc aaggccgacg ggcgcaccgt   19740
ccgcaccgtc gtgcacaccg ccgccaccat cgagctgcac accctggacg ccaccaccct   19800
ggcggacttc gaccgggtgc tgcacgccaa ggtcaccggc gcccaggtcc tcgccgaact   19860
gctcgacgac gaagagctgg acgacttcgt cctgtactcc tccaccgccg gcatgtgggg   19920
cagcggcgcc cacgccgcct acgtcgccgg caacgcctac ctcgccgcgc tcgccgagca   19980
ccgccgggcc aacggactgc ccgccctgtc gctgtcctgg ggcatctggg ccgacgacct   20040
caaactgggc cgggtcgatc cccagatgat ccggcgcagc ggcctggagt tcatggaccc   20100
gcagctggcc ctgagcggcc tgcagcgggc gctggacgac aacgagaacg tgctcgcggt   20160
cgccgacgtg gactgggaga cctaccaccc cgtctacacc tccggccgac ccacccccgct   20220
cttcgacgag gtgccggagg tccgccggct caccgcggcc gccgagcaga gcgccgggac   20280
cgtcgccgag ggcgagttcg ccgccgcgct gcgcgccctg tccgacgccg agcagcagcg   20340
caccctgctg gagaccgtcc gcaccgaggc ggcgtccgtc ctcgggctgt cctccgccga   20400
ggacctcacc gaccagcggg ccttccgcga cgtcggcttc gactcgctga ccgccgtcgg   20460
cctgcgcaac cggctcgcct ccgtcaccgg cctgacgctg ccctcgacga tggtcttcga   20520
ctaccccaac ccggccgcgc tcgccgccta tctgcacggc gagctggccg gcgcccggtc   20580
cgccgccgcc ggcgccgccg ccgtcccgac cggcgccccc gacgccgacg acccgatcgc   20640
gatcgtcggc atgagctgcc gctaccccgg cggggtcggc tccgccgagg acctgtggcg   20700
gatcgccctg gacgaggtcg acgcgatctc cggcttcccc gccgaccgcg gctgggacgc   20760
cgagggcctc tacgacccgg accccgaccg gcccggccgc acctactccg tccagggcgg   20820
attcctgcgc gacgtcgccg agttcgaccc gggcttcttc gggatctcgc cgcgcgaggc   20880
gctgtcgatg gacccgcagc agcggctcct gctggagacc gcctgggagg cgttcgagca   20940
cgccggcatc gacccggtcg gccagcgcgg cagccgcacc ggcaccttcg tcggcgccag   21000
```

```
ctaccaggac tacgcctccg gcgtgcccaa cagcgagggc tccgaaggcc acatgatcac    21060
cggcacgctc tccagtgtgc tgtccggccg ggtgtcctac ctcttcggct tcgagggccc    21120
cgccgtcacg ctcgacaccg cctgctcctc ctccctggtc gccatgcacc tggcctgcca    21180
gtccctgcgc aacggggaga gctcgctggc cctggccggc ggcgtcagca tcatgtccac    21240
cccgatgtcg ttcgtcggct tcagccggca gcgcgccctc gccgaggacg gccgctgcaa    21300
ggcgtacgcg gacggcgccg acgggatgac cctcgccgag ggcgtcggcc tggtgctgct    21360
ggagcggctg tccgacgccc gcgccaacgg gcaccaggtg ctcgccgtga tccgcggctc    21420
cgccgtcaac caggacggcg cctccaacgg cctgaccgca cccaacggcc cgtcccagca    21480
gcgcgtcatc cgccaggcgc tggccaactc cgcggtggcg cccggcgaca tcgacgtcct    21540
ggagggccac ggcaccggta ccgccctcgg cgaccccatc gaggcgcagg ccctgctcgc    21600
cacctacggc caggaccgcg cccccgaacg gccgctgctg ctcggctcgg tgaagtccaa    21660
catcggccac acccagatgg catccggcgt cgccagcgtc atcaagctcg tccgcgccct    21720
ccaggaaggc gtggtgccca agtccctgca catcgaccgg ccctccaccc acgtcgactg    21780
gtcctcgggc gccatcgggc tgctcaccga acgcaccccg tggcccgaga ccggccggcc    21840
gcgccgcgcc gccgtctcct ccttcggcat cagcggcacc aacgtccaca ccatcctcga    21900
acaggccccc gcggacgagg cgcccacgcc cgccgacccg ccgcgggacg gcctggtgcc    21960
ggtcctgctc tccggccgcg gcgaggccgc gctgcgcgcc caggccgccc gcctgctcgc    22020
cttcgtcgag gagcggcccg aggcccacct caccgacctc gcccactccc tcgccacctc    22080
gcgcgccgcg ctggaacgcc gcgccgccgt catcgccgcc gaccgcgaca ccctgacccg    22140
cggcctgcgc gccctgtccg acggccggcc cgaccccggc ctggtccagg gcaccgcggg    22200
acgcggccgg accgccttcc tgttcaccgg acagggcagc cagcgccccg gcatgggccg    22260
cgaactccac gaccgctacc cggtgttcgc cgacgcgctg gacgaggtgc tggcccggct    22320
cgacgacgga ccggaccggc cgctgcgcga ggtgctgttc gccgcgcccg actccgccga    22380
ggccgcgctc ctggaccgga ccggctacgc ccagcccgcg ctgttcgccg tcgaggtcgc    22440
gctgttccgc ctgctgacgt cctggggcct gaccccggac tacctggccg gccactccgt    22500
cggcgaactc gccgccgcgc acgtcgccgg cgtgctgtcg ctggacgacg cctgcactct    22560
ggtcgccgcc cgcggccggc tcatgcaggc gctgcccgag ggcggcgcga tggtcgccct    22620
ggaggccgcg gaggacgagg tcctgccgct cctggagggg ctcaccgacc gggtgtccgt    22680
cgccgccgtc aacgggccgc ggtccgtggt cgtcgccggc gtcgaggagg acgtgctcct    22740
cctcgccgac ctcttcgccg ccgacgggcg ccggaccaag cggctgcggg tgagccatgc    22800
cttccactcg ccgctgatgg acgccatgct cgacgacttc gccgccgtcg cccgcgggct    22860
gacctaccac ccgccgacga tcccgttcgt gtcgaacgtc agcggcggcc tggccaccgc    22920
cgaacaggtc cgcacgcccg actactgggt cgggcacgtc cgcgccgcgg tccgcttcgc    22980
cgacggcatc gactggctcg ccaccagggg cgacgtccac accttcctgg agctcggccc    23040
ggacggcgtg ctcagcgcca tggcccggga gagcctcacc gacccgtccc gcacggcact    23100
gctgccgacc ctgcgcggcg accggcccga ggaacctgcc ctggtcaccg cggtcgccgc    23160
ggcccacgcg cacggcgccc gcgtcgactg gagcgggtac ttcgccgacc acggcgcgcg    23220
ccggaccacg ctgccgacct acgcgttcca acgcgagcgg tactggcccg acaccaccgc    23280
cgccacgagc gcccacacgc ccggatccgc cctcgacgcc gagttctggg ccgccgtcga    23340
gcgggacgac gtcgccgccc tcgccgcctc cctggacctg gacgacgcca ccgtcaccgc    23400
gatggtcccc gcgctcaccg cctggcgccg gcgccgcggc gagcagaccg aactggactc    23460
ctggcgctac cgcgtcacct ggaagccgcg cggcggcgcc accgcacccg ccgccctcac    23520
cggccgctgg ctcgtgctcg tcccgcacga ccaccaggac cgtcaggacg acgcgaccgc    23580
ggcctgggca gccgacgtcg agaccgccct gggcaccacc accgtccggc tgacggtcac    23640
caccaccgac cgcgccgcgc tggccgcccg gatcaccgaa gccgccggcg accagggccc    23700
gttcagcggt gtgctgtccc tgctgccgct cgccaccggc gacgccggcc accccggtgc    23760
gcccgccgcc ctcaccctca ccaccaccgc cgtccaggcc ctcggcgacg ccggcatcga    23820
cgcgccgctg tggaacgtca cccgcggggc cgtggccgtc ggccgcgccg aacaggtcac    23880
cgcgcccgaa caggccgccg tctggggcct gggccgcgcc gtcgccctgg aactgccggc    23940
ccggttcggc ggcaccctcg acctgcccgc caccctggac ggccaggccg cccgccggtt    24000
gcgcgcggtg ctcgcggcta ccgacggcga ggacgcggtg ccctgcggc cctccggcgt    24060
cttcctccgc cgcctggccc acgccccggc cggccccgac accgcccgca ccgccttcga    24120
cccggcggcc ggcaccgtcc tgatcaccgg cggcaccggc ggcatcggcg gccacgtcgc    24180
ccgccgcctg gcccgcgacg gcgccaccca cctgctgctc accagccgcc gcggcccggc    24240
cgccccggc gccgacgcgc tccgcgccga actggaggaa ctgggcgccc gggtcaccct    24300
cgccgcctgc gacgccgccg accgcgacgc gctggccgcg ctcctcgccg aactgcccga    24360
cgacgcccg ctgtgcgcgg tgttccacac cgccggcgtc gtcgaggacc acgtcgtgga    24420
cgcgctcaca ccggagaact tcgccgcggt gctgcgcgcc aagaccgtcg ccgcccacca    24480
cctgcacgag ctgaccgccg acctggacct cgccgctttc gtgctgttct cctccacggc    24540
cggcgtcctc ggcgccgccg gacagggcaa ctacgccgcg ccaacgccc acctggacgc    24600
```

```
cctcgccgaa caccgccgct cccacggcct gaccgcgctg tccgtcgcct ggggcccgtg   24660
ggccggctcc ggcatggtcg ccgacgccgc cgaactcacc gaccgggtac ggcgcggcgg   24720
cttcgaaccg ctcgcccccg aaccggccgt gcgcgccctg ctgcgcgcca tcgagaacga   24780
cgacaccacc gtcgcgctcg ccgacatcga ctgggagcgc ttccagcgcg ccttcgccgc   24840
ggtccgcccg ctgccgttcg tcgccgacct ccccgagacc ggccgggcca cccccgcgac   24900
cgccaccggc gccgccaccg gcctgcggca gcaactcgcc gaactgccgg agcacgagcg   24960
cccggcagcg gtcctggacc tgctgcgtac ccaggtcgcc gccgtcctcg gccacgccga   25020
cccgcgcacc gtcgaggacg accacgcctt ccgcgacctg ggcttcgact cgctgaccat   25080
cctggaactg cgcaacgccc tcaacgccgc caccggcctg agcctgcccg ccaccctggt   25140
ctacgacctg cccacccgc gcgagatggc ggacttcctg ctcgccgaac tcctcggcac   25200
cctgcccacc gacaccgccg cgaccgtcgc cagcacggcc tcccccaagc tctcagcttc   25260
gttcgagcag ggcggtaccc ccttcgacga cccgatcgcc gtcatcggca tcggctgccg   25320
cttccccggc ggcgtcacca ccccggagga gctctggcag ctcctcgacg agggccgcga   25380
cggcatcagc cgcttccccg acgaccgcgg ctgggacctc gccgcgctgg gcgccggcgc   25440
ctccgacacc ctggagggcg gcttcctgac cggcgtcgcc gacttcgacg cccggttctt   25500
cggcatctcg ccccgcgagg cgctggccat ggacccccag cagcggctgc tgctggagac   25560
cacctgggag gcgctggagc gggccggcat cgacccgacc acgctgcgcg gctccaccac   25620
cggcgtcttc gtcggcacca acggccagga ctacccgacg ctgttgcgcc gctccgcctc   25680
ggacgtggcc ggctacgtcg ccaccggcaa caccgccagc gtgatgtccg gccgcctgtc   25740
ctacgcgctc ggcctcgaag gcccggccgt caccatcgac accgcctgct cctcctcgct   25800
cgtcgccctg cactgggccg gccgggcgct gcgcgccggc gagtgcgacc tcgtggtggc   25860
cggcggcgtc tcggtcatgg ccagcccgga ctccttcgtc gagttctcca cccagggcgg   25920
cctggcaccc gacgggcgct gcaaggcgtt ctccgacgcc gccgacggca ccgcctggtc   25980
cgaaggcgtc ggcatcctcg tcctggaacg cctctccgcc gcccgccgca acggccacca   26040
ggtcctcggc ctgatccgcg gcaccgccgt caaccaggac ggcgcgtcca acggcctgac   26100
cgcgcccaac ggcctctccc agcagcgcgt catcgcccag gcactcgccg acgcccgcct   26160
gcgccccgcc gacatcgacg cgatcgaggc gcacggcacc ggcaccaccc tcggcgaccc   26220
gatcgaggcc cgcgccctga tcaccgccta cggccgggac cgggacgccg aacggccgct   26280
gctgctgggc accgtcaagt ccaacatcgg gcacacccag gccgccgccg gtgccgccgg   26340
cgtcatcaag atgctgatgg cgatgcgcca cggcaccctg cccaggacgc tgcacgtggg   26400
caccccgtcc agccacgtcg actggagcgg cggcaccgtc gcgctcctcg acgacgcgcg   26460
gccctggcca cggaccgggc agccgcggcg cgccggcgtc tccgccttcg gcgtcagcgg   26520
caccaacgcc cacgtcgtcg tcgagcaggc cccggaaacc gaagcccccg ccgccccggc   26580
cgccgagccg cgcgccggag ccacgcccac cgtcgtcccc tgggtcgtct ccggacgcag   26640
ccgggaagcg ctccaggcgc agctggaccg gctcaccgcg cacaccgccg cccacccccg   26700
gcgctcggcg gcggacgtcg gccgctcgct ggccaccgac cgcacgctct ccccgcaccg   26760
cgccgtgctg ctcgccggcc cggacggggt gcgcgaggcc gcccgcgccg ccgcgccccg   26820
cacccccggc cgcaccgcgt tcctgttctc cggacagggc gcccagcacg ccctgatggg   26880
ccacgacctg taccagcgct tcccggtcta cgccgacgca ctggacaccg tcctcgccca   26940
gttcgacacc gtgctggacg tcccgctgcg cgccgcgctg ttcgccgcgc cgggcacccc   27000
cgaggccgcg ctcctggacc agaccggctt cacccagccc gcgctgttcg ccgtcgaggt   27060
cgcactgttc cggctcgccg agtcctggcg gctgacgccg gacttcgtcg ccggccactc   27120
catcggcgag atcgccgccg cgcacgtcgc cggggtgttc tccctggagg acgcctgcac   27180
gctggtcgcc gcccgcgcct ccctcatgca gcaactgccg cgggacggtg cgatggtggc   27240
cctggaagcc accgaggacg aggtcgcgcc gctgctcacc gacggcgtcg cactcgccgc   27300
ggtcaacggc ccccgctcgg tggtcgtcgc gggcgccgag gacgccgtcc gcgcggtcgc   27360
cgaccggctc gccgccgacg gccgccgcac ccgccggctg acggtcagcc acgccttcca   27420
ctcgccgctg atggacccga tgctcaccga cttcgcccgg gtcgcggagg cctgacccta   27480
ccacgagccg cgcatccccc tcgtctccac cctcctcggc gccccggccg cgccgaact   27540
gcgcacccc gactactggg tgcggcacgt ccgcgagacc gtgcggttcg ccgacggcgt   27600
gcgcgccctg cacgacgccg cgccggcac cttcgtggag atcggcccgg acggcgtgct   27660
caccgccctg acccagcaga ccctcgacac cgtcgaggcc ggcgcgcccg ccgtcgtcgt   27720
gccgctccag cgccgcgacc gcgccggcga cctcgccctc ctggagggcc tggccaccct   27780
gcacacccac ggcaccggcc cgtcctggcc cgcctacttc gaggccaccg gcggccaccg   27840
gaccgatctg cccacctacg ccttccagcg ggagcggtac tggcccgaac tcggcgcacc   27900
cgtcgccacc gccccgcagg acccggcggc ctggcgctac cacgagacct gggccccgct   27960
gccggcaccc gaggcggccg cgccggccgg ccgcgccctg gtcctcgtcc cggccgggaa   28020
ccgcgacacc gcgtggatga cggccgtcgc cgacgcgctc ggcgccgaca ccgtcacggc   28080
cgaacccgac gccctggccg agcagttgac ggccgccggc gacacaccct ggcgcgtcgt   28140
ggtgtcgctg ctcgccgccg cgtccgaggg gctgcccgcg gacggcgcct ggccgccgc   28200
```

```
cctcctcgcc accctggacg aggccggcgt gcacgcgccc ctgtggtgcg tcacccgcgg   28260
cgccgtcgcg gtcgcggggg aggccccgac cgccgtcggc caggccgccc tgtggggcct   28320
gggccgggtc gccgcgctgg accacccgga ccgcttcggc ggcctggccg acctgcccgc   28380
cgacaccgac gcgcacgccg ccgggctgct cgccgcgcac ctggccgcgc cgggcaccga   28440
ggccgagatc gcggtccgcg ccaccggcgt ccacgcccgt cgcctggtcc gtacgccggc   28500
cgccgccgac ggtgccacct ggctgccgac cggcaccgtc ctggttgtcg gcggcaccgg   28560
cggcaccggc accatgggcg gccgggccgc ccgctggctg gtccgcgagg cgcccgcca   28620
cctcgtcctg accgcccccg acggcaccac gaccgccgcg gacaccgagg ccctgacggc   28680
cgaactggcc gcgctcggcg cccggatcac cgtcgtggac cacgaccca ccgccccgga   28740
cggcttcgcc gcgctcctcg acggactgcc cgacgacacc ccgctcaccg cggtcgtgta   28800
cgcgccggag gccgacgccg cccccggcac cgcggccgag ctgtccgccg cactcgcccc   28860
cgtcaccgcc ctaggcgccg ccctcaccgg ccggccgctg gacgccttcg tcctcttcgg   28920
ctccatcgcc gggctctggg gcgtgcgcgg ccgggccgcc gaggccgcgt ccggcgccta   28980
cctcgacgcc ttcgcccgcg cctgccgcga ccgcggcacc ccggcactgg ccgtcgcctg   29040
gggcgcctgg gccgacctgg tcggcccgtc cctcgccgcg cacctgcgga tgaacggcct   29100
gccggtgatg gacgcggaca ccgcactgac cgccctcagc cgggccgtcg ccgacgggtc   29160
cgccgccgag gcggtcgccg acgtccgctg ggagaccttc gcgcccctcc accacgaggc   29220
ccgccgcacc gccctgttcg acgccctgcc cgaggcccgc ggcgcgctcg cggaggccgc   29280
ccgggaccgc gccgaccgga agaccgccgc cggcgactac ggccggtggc tcgccgagca   29340
gcccgccgcg gaccacgacg ccatcctgct ggcactggtc accgagaagg ccgcgaccgt   29400
cctcggccac gccgaccacg acctgctcga acccgacctg cccttccgcg acctgggctt   29460
cgactcgctg accgcggtcg acctgcgcaa ccagctcacc gcggaaaccg gcctcaccct   29520
gcccgccacc ctcgtcttcg accaccccaa cccggccgcc ctcgccgccc acctgcgcgc   29580
ccaactcctc ggcgaggcga gcgactccgc cgcaccggtg gccgcccccg tcgccctcgg   29640
tgccgacgac gacgcgatcg tcatcgtcgg catggcctgc cgctaccccg gcggggtcac   29700
ctcgcccgag gacctgtggc agctggtcgg cgacgaggtc gacgcggtcg gcgacttccc   29760
gaccgaccgc ggctgggacc tggccgcgct cgccggcgac ggaccgggcc gcagtgccac   29820
cgcccagggc ggattcctct acgacgccac cgacttcgac cccggcctgt tcggcatctc   29880
gccgcgcgag gccctggtga tggacccgca gcagcggatc ctgctcgaaa cgtcctggga   29940
ggccctggag cgggccggca tcgacccggc gacgctgcgc ggcagcggca ccaccggcgt   30000
cttcgtcggc ggcggctccg gcgactaccg gccgccggag gaggccgggc agtggcagac   30060
cgcccagtcc gccagcctgc tctccggtcg cctcgcctac accttcggca tccagggccc   30120
caccgtgtcg gtcgacaccg cctgctcctc gtcgctggtc gcgctgcacc tggccgcgca   30180
ggccctgcgc gccggcgaat gctcgatcgc gctggccggc ggcgtcaccg tgatggccac   30240
cccggtgggc ttcgtcgagt tcagcgccca gggcgccctg tcgccggacg ccgctgccg   30300
cgccttctcc gacgacgcca acggcaccgg ctggtccgaa ggcgtgggca tgctcgtcgt   30360
cgaacggctc tccgacgccc gccgcaacgg ccaccgcgtc ctcgccgtgc tccgcggctc   30420
cgccatcaac caggacggcg cgtccaacgg cctgaccgcc cccagcggcc ccgcccagca   30480
gcgcgtcatc cgccaggccc tcgccaacgc ccgactgcgc cccgccgaca tcgacgccgt   30540
cgaggcccac ggcaccggca ccaggctcgg cgaccccatc gaggcccagg ccctgctcgc   30600
cacctacggc caggaccgcg agcggcccgt gctgctcggc tcgctcaagt ccaacatcgg   30660
ccacacccag gccgcctccg gcgtcggcgg cgtcatcaag atggtcctcg ccatgcagca   30720
cggcgaactg ccgcgctccc tgtacgccga gaacccctcg tcgcacgtgg actggaccgc   30780
cggccgcgcc cacctgctca ccgccaggac cccgtggccc gactccggtc ggccgcgccg   30840
cgccgccgtc tcctccttcg gcgccagcgg caccaacgcc cacgccatcc tggagcagcc   30900
gccgcgcgag gaactccccg cgcgccccgc ggacgacggc gccccgctgc cgttcttgct   30960
ctccggccgc tcgcagaacg ccctgcgcgc ccaggcccgc cgactcctgg cccgcctcac   31020
cgcccacccc gacacccggg ccgccgacct ggcgtactcc ctggcgacca cccgggccgc   31080
cttcgagcac cgggccgcga tcaccgccac cgaccacgac ggcctccgca ccggcctgac   31140
cgccgtcgcc gagggcacca ccgccccgca caccgccgaa caccacctcc agggcaccgg   31200
aaagcgcgcc gtgctcttct ccggccaggg ctcccagcgc ctgggcatgg ccgcgaact   31260
gcacgagcgc cacccggtgt cgccgaggc gttcgactcc gtactggccc gcctcgacga   31320
ccggctcgac acccccctgc gggacgtcgt ctggggcacc gacgaggagg cgctgcacgc   31380
caccgggaac acccagcccg ccctgttcgc cgtcgaagtc gcgctctacc gcctgatcga   31440
atcctggggc gtgcggcccg acttcgtggc cggccactcc gtcggcgagc tcgccgcggc   31500
ccacgtcgcc ggggtgctct ccctggacga cgcctgccgc ctggtcgccg cccgcgccgc   31560
cctcatgcag cgcctcccgg ccggcggcgc catgatcgcc gtcgaggcca ccgaggacga   31620
ggtcacccc ctcctcaccg acggcgtgtc cctcgccgcg gtcaacggac cgaccgccgt   31680
ggtcctctcc ggcgcggggcg acgccgtgac cgccctgggc caggcgctgg ccgaacgggg   31740
ccaccgcacc accggctgc gggtcagcca cgccttccac tcgcacctca tggacccgat   31800
```

```
gctggcggac ttccgcaccg tcgccgaggg cctggaatac cacccgccgc gcatccccgt    31860
ggtctccaac ctcaccgggg acgtcgccga cgcggccgac ctgtgctccg ccgactactg    31920
ggtgcgccac gtccgcggca ccgtacggtt cgccgacggc gtgcgcacca tggccgaccg    31980
cggcgtgcac ctcttcctcg aactcggccc ggacgccgtg ctgtcggcca tggcccgcca    32040
gtgcgcaccg gacgccgtcg tcgtcccggc cctgcgccgc aaccgcgacg aggacgagac    32100
gctggtcggc gccgtcgcgc gactgcacgt ccacggcgcg ggtccgcgct gggacgcgta    32160
cttcgccggc cgcggcgccc agtggctgga ccttccgacg tacccc ttcc agcgcggccg    32220
cttctggccg gagtccct tc cgggcgccgc atcggccgcc ccggcagccg gacagccggc    32280
cgagaccgac gcggccttct gggacgccgt cgcacaggag gacttcaccg cattggaatc    32340
cgtactcgac gtcgagagcg acgcactgtc caaggtgctg ccggccctga tggactggcg    32400
cagccgccag gccgacgagt cccaactggc aggctggcgc caccgcatcg tctggaagcg    32460
gctcaccggc gccgccctgg cacaccgcaa ggcgctcagc ggcacctggc tcgcggtggt    32520
ccccgagggc ttcgccgacg acccctgggt gaccaccacc ctggacggcc tcggtaccca    32580
cctcgtgcat ctggaggtcg cggaggccga ccgggccgcg ctggccgacg cgatcgcggc    32640
ccgcaccgcc gacggcaccc gcttcggcgg cgtaatctcc ctgctggccc tgcgcgagga    32700
gctcaccggc gcggtgcccg aggggaccgc cctgaccacc accctcctcc aggccctcgg    32760
cgacgccggc gtcgacgcac cgctgtggtg cgtcacccgc agcgccgtct ccgccggccg    32820
caccgaccgg ccgcaccgac cgctccaagg cgccgtctgg ggcctgggcc gggtcgcggc    32880
ccttgagtac ccgcagcgct ggggcggcct ggtggacctg ccggaggagc ccgacgagcg    32940
gtccgcggcc ggcctcgccg ccgtcctggc cggtctggac ggcgaggacc aggtcgccgt    33000
gcgcggcacc gcggtgctcg cccgccgcct ggtgccggct cccggccgca agccgtcccg    33060
gccctggcac ccgtccggca ccgtcctggt caccggcggc accggcgccc tcggcgcgca    33120
cgtcgcccgc cgcctggcca aggacggcgc ccagcacctc gtcctgctca gccgccgcgg    33180
cccggacgct cccggtgcgg cggaactgcg cgcggaactg gacgcgttgg gcaccgacgt    33240
cacggtcgcc gcctgcgacg tcgccgaccg cgaccagctg acggccgtcc tggacgcgct    33300
gcccgccgac cggccgctga ccggtgtggt gcacaccgcc ggcgtcctcg acgacggcgt    33360
actggaccgg ctcacccccg agcggttcca ggaggtgttc cgcgccaagg tcacctcggc    33420
cctgctgctg gacgagctga cccgcgaccg cgagctggcc gcgttcgtcc tcttctcctc    33480
cgcctccgcc gcggtcggca acccgggcca ggccaactac gccgctgcca acgccgtcct    33540
ggacgcgctc gccgaacagc gccgggtgct cggcctgccc gccacctcgg tctcctgggg    33600
tgcctgggga ggcggcggca tggccgacgc cgacggcgcg gacgaggccg cccggcgcgc    33660
cggcgtcggc gccatggacc cgcacctcgc cgtggaagcc ctgctgcgcc tggtcgccga    33720
gaaggagccg accgcggtgg tcgccgaggt ggccctggac cggttcgccg gcgccttcgg    33780
cggcagccga cccagcgccc tgctgcggga gttccccggc taccgcgagg cgctcgccgc    33840
ccaggcggag caggccgcgg acggcggcgg gctggccgcc cgactggccg cgctgccgcc    33900
cgcccgccgc ctggacaccg ttgtggacct ggtgcgcacc cgcgccgcgc aggtgctcgg    33960
ctaccccgac accgaagcgg tcgccgccga acggtccttc cgcgacctgg gtgtcgactc    34020
gctcggcgcc gtcgagctgc gcaaccaact gagcgcggcc accggcctga acctgccggc    34080
gacgctggtg ttcgaccacc cgaccccct ggtcctgggg gagcacatcc tcggcgggct    34140
cttcccggac gagcccgccg ggtccgacga cgagacggag atccgggccc tgctggcctc    34200
cgtcccgctc gaccaactgc gggagatcgg ggtcctggag ccctgctcc agctcgccgg    34260
acgcggcggc cgggccgcgg acggcgacga cggcgagtcc gtcgactcga tgacagtggc    34320
agacctggtg cgggccgcgc tcaacggcca gtccgacctg tagcgcgatt gatggagcag    34380
acgatgaacg cgcccgagaa ccccgagacc cccgagaaca acgtagtcgc cgcactccgc    34440
gccgcggtca aggagaccga ccggctccgg cggcagaacc ggatgctggt cgcggcggcc    34500
aaggaaccga tcgccgtggt cggcatggcc tgccgcttcc ccggcgccgt cgactccccg    34560
gaagcgctgt gggagatggt cgccaccggc accgacgtga tctccggatt ccccgacgac    34620
cgcggctggg acctggaggc gctgcgcaac agcggcaccg acgcccgcga caccgacgtc    34680
agccagcgcg gcggattcct ggactgcatc gccgacttcg accccggctt cttcgggatc    34740
tcaccgcgcg aggcggtcac catggacccg caacagcggc tcctgctgac caccgcctgg    34800
gaggccgtcg agcgggccgg catcgacgcc accacgctgc gcgccacccg caccggcgcg    34860
ttcatcggca ccaacggcca ggactacgcc tacctgctcg tccgctccct ggacgacgcc    34920
accggcgacg tcggcaccgg catcgccgcc agcgccgcct ccggtcggct ctcctacacc    34980
ctcggcctcg aaggccccgc gctcaccgtc gacaccgcct gctcctcgtc gctggtcgcc    35040
ctgcacctgg ccgtgcaggc gctgcgcaac ggcgagtgcg gcatggcgct ggccggcggc    35100
gtcaacgtga tggccacacc gggctcgctg gtcgagttca gccgccaggg cgggctggcc    35160
cgggacggcc gctgcaaggc gttcgcggac gccgccgacg gcaccggctg gtccgagggc    35220
gccggcgtgc tgctgctgga acggctctcc gacgcccagc gcaacggcca cccggtgctc    35280
gccgtggtcc gcggctccgc cgtcaaccag gacggcgcct ccaacggctt caccgccccc    35340
aacggcccct cccagcagcg cgtcatccgc caggccctcg ccaacgccgg cctggccacc    35400
```

```
ggcgacatcg acgcggtcga ggcgcacggg accggcaccc cgctcggcga ccccatcgag    35460
gcgcagagca tcctcgccac ctacggccag gaccgcgccc acccggtgct gctcggctcg    35520
atcaagtcga acatgggcca cacccaggcc gcgtccggcg tcgccggcgt gatcaagatg    35580
atcatggcga tgcggcacgg cgtcctcccg cggaccctgc acgtcgaccg gccctccacc    35640
cacgtcgact ggaccaccgg cagcgtcgaa ctcctcaccg acgcccaccc gtggcccgag    35700
actggcaggc cgcgccgcac cggcatctcc tccttcggcg tcagcggcac caacgcccac    35760
gtcatcgtcg aacaggcccc cgacacccccc gccgaggcgg ctgacgacac tccgccccgc    35820
accccgcgga ccctgccgtg gctgctctcc gcccgcaccg gcgccgccct cgcgcgaccag    35880
gccaccgcgc tgctcgacca cctcgaccgc cccgacggcg accgcgggcc caccgccctg    35940
gacaccgcgt tctccctcgc caccacccgc gccgccctgg aacaccggct cgccgtcgtc    36000
accggcaccg acggcaccgc cggacgggac gccctgaccg cctggctggc gcacggcacc    36060
gcccccgacg cccacgaagg acacgccgcc ggacgcaccc gctgcgcggc cctcttctcc    36120
ggccagggcg cccagcgcct gggcatgggc cgcgaactcc acgcccgttt cccggtgttc    36180
gcacgggccc tcgacaccgc cgtggacctg ctcgacgccg aactgggcgg caccctgcgg    36240
gaggtgatct ggggcaccga cgacgcgccg ctcaacgaga ccggcttcac ccagcccgcc    36300
ctgttcgccg tcgaggtcgc cctctaccgc ctgatcgaat cctggggcgt cgccccggac    36360
ttcgtcgccg gccactccat cggcgagatc gccgccgcgc acgtcgccgg ggtgttctcc    36420
ctggaggacg cctgcacgct ggtggccgcc cgcgccgggc tgatgcaggc gctgccgcgc    36480
ggcggggcga tggtcgccgt cgaggccacc gaggacgagg tcagcccgct gctcaccgac    36540
ggcgtcgcga tcgccgcgat caacggcccc acctcgctcg tcgtctccgg cgacgagacc    36600
gccacccctcg ccgtcgccgc ccgactcgcc gaacagggcc gccgcaccac ccggctgcgg    36660
gtcagccacg ccttccactc gccgctgatg gacccgatgc tcgcggagtt ccgcgcggtc    36720
gccgagggcc tgtcctacgg cgaaccgcag atcccggtgg tctccaacct caccggcgcg    36780
gtcgccgacg gcaccctgct cggcactgcc gactactggg tccggcacgt ccgcgaggcg    36840
gtccgcttcg ccgacggcat ccgcgccctc accgacgccg gcgtcggcgc cttcctcgaa    36900
ctcggcccgg acggcacgct cgccgccctg gcccagcagt ccgcccccga cgccgtctcc    36960
gtccccgtcc tgcgcaagga ccgggacgag gagcccgccg cggtcgccgc actggcccgg    37020
ctgcacaccg ccggcgtccc ggtggactgg acggcgttct acgccggcac cggcgcccac    37080
cgcaccgacc tgccgaccta cgccttccag tacgagcgct actggcccaa ggccacctac    37140
cggcccgccg acgccaccgg cctcggcctg accgccgccg accacccgct gctcggcgcc    37200
gccatgtccg tcgccgggtc cgacgagctc ctgctcaccg gcaccctgtc gctcgccacc    37260
cacccctggc tcgccgacca cgtcgtcggc ggcatggtct tcttccccgg caccggcttc    37320
ctggaactgg cggtccgcgc cgccgaccag gtcggctgcg accgggtcga ggaactcatg    37380
ctcgccgcgc cgctgatcct gccccgccacc ggcaccgtcc agatgcagat cgcggtcggc    37440
gccgcggacg acgacggcgg ccgcgacctg cgcttcttca cccggcccgg ggacgacccg    37500
gacgccgcct gggcccagca cgccaccggc cggatcaccg agggcgagcg cgtcctcgcc    37560
ctcgacacca ccacctggcc gccccgcgac gcggaacccg tcgacatcga cggcctctac    37620
gaccgctacc gcgccaacgg actcgactac gggcccgtct tccgcggcct gcgcgccgta    37680
tggcgccgcg acaccgagat ctacgccgag gtcgccctgc ccgaaggcac cgccgacgcc    37740
gacgccttcg gcctgcaccc ggccctcttc gacgccgtcc tgcacagcac cctcttcgcc    37800
tccgccgacg gcgacgaccg cagcctcctg ccgttcgcct ggaacggcgt gtccctgcac    37860
gccgcgggcg cggacgcgct gcgcgtccgg atcaccagct gcggccccga cgccgtggag    37920
atcaccgccg tcgacccgca gggccgcccc gtcgtctccg tcgaatcgct gacgctgcgc    37980
gccgccggcc ccgacgccgg caccgccgac caccgtgccg acgcgggctc cctcttccgc    38040
atggactgga cgccccgcac cgtccacgcc ccggccaccc ccgccacctg ggccgtcctc    38100
ggtaccgacc cgatcggcct gaccgaggcg ctcaccgccg ccggccccga caccgtcacg    38160
ggactccgcg acggtgtcga cgccctcggc gaactcaccg ccggcgacga ccggccggtg    38220
cccgacgtgg tcgcggtacc gctgcgcggc gccaccgacc acgggccggc cggtgcccac    38280
gacctgaccc gcaccgtcct ggccctgctc caggaatggc tggccgagga gcgcttcgcc    38340
cgctcccggc tgctgctcgt cacccgcggc gcggtcgccg acggcgagcg cggcccgctc    38400
gacctggccg ccgccccggt ctggggcctg gtgcgctccg cccagtccga gaaccccggc    38460
cgactgctgc tcgtcgacct cgacgacacc gccgagtccg ccgcccaact cccgttgctg    38520
ccggccctcc tggacgccga cgagccgcag gccgtggtcc gcgagggcac cgtccgggtc    38580
ggccggctcg cccgcctgga ctccggccgc ggcctcgtcc cgccgcccgg caccccctgg    38640
cgcctgggca gccgggccaa gggcagcctc gacggcctcg ccctgctgcc ccaccccgag    38700
gcccgacgac ccctcaccgg ccacgaggtc cgcgtcggca tccgcgccgc gggcctgaac    38760
ttccgtgacg tgctcaacgc gttggggatg tatcccgggg atgcggggct gttcggttcg    38820
gaggcggccg gtgtggtcgt cgaggtcgga ccggaggtca ccggcctggc acccggcgac    38880
cgggtcatgg gcatgctctt cggcggcttc ggaccgctcg gcatcgccga cgcccggctg    38940
ctcacccccg tcccggccga ctggtcctgg gagacgggtg cgtcggtgcc gttggtgttc    39000
```

```
ctcaccgcgt actacgccct gaaggagttg ggtggtctgc gggcggggga gaaggtgctg    39060
gtgcatgccg gtgccggtgg tgtcggtatg gcggcgatcc agatcgcccg gcatgtcggt    39120
gccgaggtgt tcgccacggc cagtgagggc aagtgggacg tgctgcgctc cctgggcgtg    39180
gccgacgacc acatcgcctc ctcccgcacc ctcgacttcg aggcggcctt cgccgaagtc    39240
gccggcgacc gcggcctgga cgtcgtactg aacgcgctgt ccggcgagtt cgtcgacgcc    39300
tcgatgcggc tgctcggcga cggcggccgg ttcctggaga tgggcaagac cgacatccgc    39360
gccgcggact ccgttcccga cggcctctcc taccactcct tcgacctcgg catggtcgat    39420
ccggaacaca tccagcggat gctgctcgac ctcgtcgagc tgttcgaccg cggcgcgctg    39480
gccgcgttgc cggtccgcag ctgggacgtg cgccgcgccg gcgaggcgtt ccgcttcatg    39540
agcctggccc agcacatcgg caagatcgtg ctcaccgtgc cgcaacccct cgaccccgac    39600
ggcaccgtgc tcctcaccgg cggcaccggc ggcctggccg gcctgctcgc ccgccacctg    39660
gtcaccgagc acggcgcccg ccacctgctg ctggccggcc ggcgcggccc cgacgcgccc    39720
ggcgccgccg cactccacgc cgaactgacc gccctgggcg ccgaggtcac cgtcgccgcc    39780
tgcgacgtcg ccgaccgcac cgcgctcgcc gcgctgctcg ccaccgtgcc cgccgaacac    39840
cccctcaccg cggtcgtgca caccgccggc gtcctggacg acggcaccct caccgccctg    39900
aaccccgacc gcctcgccac cgtcctacgg cccaaggtgg acgccgcctg gcacctgcac    39960
gacctcaccc gccacctcga cctggccgcg ttcgtgctct actcctccac cgccggcgtc    40020
atgggcggac cgggccaggc caactacgcg gccggcaaca ccttcctcga cgcgctcgcc    40080
gcccaccgac acgccctcgg cctgcccgcc acctcgctgg cctggggcgc ctgggagcag    40140
ggcgccggca tgaccggcgc actgaccgac cacgacctgc gccgggtcag cgacgccggc    40200
ggccaaccgc tgctcaccgc cgaacgcggc ctcgccctct acgacgccgc caccgccgcc    40260
gacgaacccc tgatcgtccc gctcggcctc accggcggtg cgctgcccgc cggggtcggc    40320
gtccccgccg tgctgcgcgg cctggtccgc accgcgggcc gccgggccag ggccggcacc    40380
gccggcgtct cccgcgccgg cctcgccgaa cgcctcgccg ccctgcccga ggaggagcgc    40440
accccctttcc tcgtcgagct ggtgcgcacc gaggccgcca ccgtcctcgg ccacggctcc    40500
accgacccgg tggacgcccg ccgcgagttc cgccaactcg gcttcgactc gctgaccgcc    40560
atcgaactgc gcaaccgact cggcaaggcc accggcctca ccctgcccgc caccctcatc    40620
ttcgactacc cgacccccga ccgcctcgcc gtccacctcc acgacgaact cctcggcgcg    40680
gacgccccgg tgaccgtcac cgccgccgca caggccgcgg acccggagca cgacccggtc    40740
gtcatcgtcg gcatgagctg ccgcttcccc ggcggcgtca gctcccccga ggagctgtgg    40800
gacctggtgg catccggcac cgacgcgatc accggcttcc ccgccgaccg cgcatgggac    40860
cgccacccgc agctcgccgg cgcccccggc gcccgcaccg gccagggcgg attcctccgc    40920
gacatcgccg acttcgacgc cgccttcttc ggcatctcgc cgcgcgaggc cctggccatg    40980
gacccgcagc agcgcatcct cctcgaagtc gcctgggagg ccgccgagcg cgccggcatc    41040
gacccgcaga ccctgcgcgg cagcgacacc ggcgtgttca tgggcgtcag cggccaggac    41100
tacgccggcc tcgtgatgcg ctcccgcgac gacatcgccg gccacgccac caccggcctc    41160
gccgtcagcg tcgtctccgg ccgcctcgcc tacgcgctcg gcctggaggg cccggccctg    41220
tccgtggaca ccgcctgctc ctcctccctg gtgtcgctgc acctggccgc ccaggcgctg    41280
cgcgcggggg agtgcaccat ggccctggcc ggcggcgtca ccgtcatgac caccgccgcc    41340
aacttcaccg gcttctcccg gatgggcggc ctcgcccagg acggccgctg caaggcgttc    41400
tccgactccg ccgacggcac cggctggtcc gagggcgccg ccgtcctggt cctggaacgc    41460
ctctccgacg cccggcgcgc cggccaccgc gtactggccg tggtgcgcgg ctcggcggtc    41520
aaccaggacg gtgcgtccaa cggtctgacg gcgcccaacg gtcccgccca gcagcgcgtc    41580
atccggcagg ccctggccaa cgcgggcctg accccgtcg acgtggacgc cgtcgaggcg    41640
cacggcaccg gcaccccgct cggcgacccc atcgaggccc aggccctgat cgccgcctac    41700
ggcaccgacc gcgaccccga acacccgctg ctgctcggct cggtgaagtc caacatcggc    41760
cacacccagt ccgcggccgg cgcggccggg ctggtcaaga tggtcatggc catgcgccac    41820
ggcatcctgc cgcagaccct gcacctcacc gaaccgtcct cgcacgtgga ctggtcggcg    41880
ggcacggtgc ggctgctcac cgagcggacc gcctggccgc ggacggatcg tccgcgtcgg    41940
gccggggtct cctcgttcgg catcagcggc accaacgccc acgtcatcct ggaacagccg    42000
cccgccgagc ccacccccgc cgccgacccg ggccggcccg cacccaccgt ggtggcctgg    42060
cccgtctccg cgcagacccc ggccgccctc gacgcccaac tggaccggtt gcgcaccgcc    42120
gccgccctgg cgccgctcga caccgcccac accctcgcca ccggccgctc gctcttcgaa    42180
caccgcgccg tcctgctcgc caccgtcggc gacccggcga ccggcgcccc cgacctgccc    42240
gaggtcgcca ggggagcggc gacgccgcac cgcaccgcgt tcctcttctc ggggcagggt    42300
gctcagcggt cggggatggg gcgtgaactg catgctgctt tcccggtgtt cgcggcggcg    42360
ttcgacgagg tggtggctgt gttggatgcg gagctgggtt ctgatgctga tggggtgtgt    42420
tcgctgcggg aggtgatgtg gggcggggg tcggagttgt tggatcgaac gcgtttcacg    42480
cagccggcgt tgttcgcggt ggaggtggcg ttgttccgtt tggtggcctc gtggggggtg    42540
gggcctgagt tcgtggcggg gcattcggtg ggtgagattg cggcggcgca tgtggccggg    42600
```

```
gtgttctcgt tggtggatgc gtgtcgtttg gtggtggcgc gggcttcgtt gatggatgcg    42660
ttgccggtgg gtggcgtgat ggttgcggtg gaggcggccg aggcggaggt ggtgccgctg    42720
ttggtcgatg gggtggcgat cgccgcggtc aacgggccgg tctcggtggt ggtctccggt    42780
gtggaggcgg ccgttgggca ggtcgtggat cagttggtgg agcggggccg gcgggtccgt    42840
cggttggcgg tcagtcatgc tttccactcg ccgttgatgg atccgatgtt ggatgccttc    42900
cgggccgtcg ccgagggcct ggagtaccac cagccgcgca tccccgtggt gtccaacgtg    42960
acgggcgagg tggccgcggc ggaggagctg tgcgcggccg actactgggt gcggcacgtc    43020
cgggcgaccg tgcggttcgc cgacggcgtc cgcaccctgg ccgagcgcgg cgccaccgcc    43080
ttcctggaga tcggccccga cggcgtactg tccgcgctcg cccgcggcgt cctgcccgcc    43140
gaggcgctcg tgacgcccac cctccgcaag gaccgcgacg aggagagcgc cctgctcgcc    43200
ggactggccc ggctgcacgt cgccggcgtg accgtcgact ggagcgccgc cctgaccggc    43260
accggcgcgc gcggcaccga cctgccgacc tacgccttcc aacgcgagcg gtactggccg    43320
gagttggccg ccgaacccgc gggcggcggc gcggatgccg cggacgcgga gttctgggcc    43380
gcggtggagc gcgcggacgc caccgcgctc gccgcccacc tggacatcga cggcgaccag    43440
ctcggcgccg tgctgcccgc actgtccgcc tggcgcaccc ggcgccgcac cacatcggcc    43500
accaacgccc tgcggcaccg ggagagttgg gaaccgctgt cgctcgccgg cacgccgcac    43560
accggcggcg tcctggtgct ggtgcccgcc gccgcgacca ccgacccctg ggtcgccgac    43620
gtcgtcgccg ccctcggccc ggacgcccgc cgggtggacg tcccggccga cggcaccgac    43680
cgggccgcgc tcgccgccct gctcaccgaa gcggccgacg acaccgcccc gaccgccgtg    43740
gtctccctgc tcgcgctcga cgagaccagc ggcgacgacg cggtaccggc cggcaccacc    43800
gccaccgccg cgctcgtcca ggccctcgcc gacaccggcg ccccggcccc gctgtgggcc    43860
ctgacccggg gcgcggtcgc cgcgctcccc gacgagcagc cgaccgcccc cgcccaggcc    43920
gccgtctggg gcctcggccg gatcgccgcc ctcgaactcc cgcgccactg gggcggactg    43980
gtcgacctgc ccgccgacct cgacgagcgc accgcacgcc gactgcccgc cgcactggcc    44040
gacgccggcg acgaggacca gctcgcgctg cgggccaccg gcgcctacgg ccgccggatc    44100
accccggcgc ccgcccccga cgacgccccc ggcaccgggt ggcagccgac cggcaccgtc    44160
ctgatcaccg gcggcaccgg cgcgctcggc cggcacaccg cccgctggct cgccgcccac    44220
ggcgccgagc acctgctgct gctcagccgc agcggccccg acgcgcccgg cgcggccgaa    44280
ctcaccaccg aactcaccgc cctcggcgcc cgcgtcaccc tcgtggcctg cgacgccgcc    44340
gaccgggagc agctgaccag ggtcctcgcc gaggtaccgc gggactgccc gctgaccggc    44400
gtggtgcaca ccgccggagt gctcgacgac ggcgtgctca ccggcctcac cccggaccgg    44460
ttcgccacgg tcttccgcgc caaggtggcc tccgccgtgc tcctcgacga gctgacccgg    44520
gacaccgacc tggccgtctt cgcgctgttc tcgtccgtcg cgggcgctgt cggcaacccc    44580
ggccaggccg gctacgccgc cgccaacgcg gtcctcgacg ccctcgccgc ccgccgccgg    44640
gcccagggcc tggccggcac ctcgatcgcc tggggtgcct gggccggcga cggcatggcg    44700
gcccgtcaca cccgccccgg cgccgaaccc gtcggcctgc tcgacccga cctcgccgta    44760
ccggccctgg cccgcgcggt gacggagccc cagcccaccc tcgtcctcgc cgacctccag    44820
cagccgcggc tgctggaatc cctgctggcg ctgcgcccca gcccgctcct gagccggctg    44880
cccgccgccc gcaccgcggc ccgcgcggtc caggaagcgg accgccgccg agccggcgcc    44940
gcggccgacc tgcgcgacca actcgccggc accgcacccg ccgaccgcca cgccgtcctc    45000
ctccgcctgg tgcggaccac ggccgccgcg gtcctcggcc acaccggtgc cgacgccatc    45060
cgggccgaca agcccttccg cgacctcggc ttcgactcgc tcaccgcggt ggaactgagc    45120
agcgctctcg ccgccgccac cggcctggcc ctcccgccca gcctcgtctt cgaccacccc    45180
tccccgcggg cgctcgccga ccacctgcgc gccgaactca ccggcgaccg gccggaatcc    45240
gcccccgcgg ccccgccggc accggtcccc gccgcggacg acgatccgat cgtcgtcgtc    45300
ggcatggcct gccgcttccc cggcggcgtc accaccccg aggagttctg gcagctgctc    45360
gccgagggcc gggacggcat cgacgcgttc cccaccgacc gcggctggga cctcgacgtg    45420
ctcggccggc gacggccagg gccacagcgc ccaccgaggt cggcggcttc ctcttacgac    45480
gcggccgcct tcgacccgg cttcttcgac atctccccgc gcgaggcgct cgccatggac    45540
ccgcagcagc ggctgctgct ggagaccgcc tgggaggccg tcgaacgcac cggcaccgac    45600
ccgaccggc tgcgcggcag ccgcaccggc gtgttcgtcg caccaacgg ccaggactac    45660
gccggcctcg tcctgcgcgc ccaggaggac gtcgagggc acgccggcac cggactggcc    45720
gccagcgtga tctccggccg cctcgcctac gccttcggct tcgagggccc cgccgtcacc    45780
gtcgacaccg cctgctcctc ctcgctcgtc gccctgcact gggccgtcca ggcgctgcgc    45840
gcggggagt gctccctggc cctggccggc ggcgtcaccg tcatgacgac ctcgacgagc    45900
ttcgccggct tcacccggca gggcggcctg gcgccggacg gcactgcaa ggcgttctcc    45960
gactccgccg acggcaccgg ctggtccgag ggtgtgggcg tcctcgtcgt cgaacgccgc    46020
tccgacgcgc tccgcaacgg ccatgagatc ctggccgtgg tgcgcggctc ggcggtcaac    46080
caggacggtg cgtccaacgg tctgacggcg cccaacggtc ccgcccagca gcgcgtcatc    46140
cggcaggccc tggccaacgc gggcctggcc cccggcgacg tggacgcggt cgaggcccac    46200
```

ggcaccggca ccgtcctcgg cgaccccatc gaggcccagg cgctgctcgc cacctacggc   46260
caggaccggc ccgccgaccg gccgttgtgg ctcggctcgg tgaagtccaa catcggccac   46320
acccaggccg ccgccggcgc cgccggcctg atgaagatgg tgctggccct ccaacacggc   46380
acgctgccgc gcaccctgca cgtcaccgag ccctcgaccc gggtcgactg gtcggccggc   46440
gcggtgcggc tgctcaccga gcggaccgtc tggccgcgga cggatcgtcc gcgtcgggcc   46500
ggggtctcct cgttcggcat cagcggcacc aacgcccacg tcatcctgga acagccgccc   46560
gccgagccca cccccacggc ccctgccgac cgccccaccc ggacgcccgc cgtcctccca   46620
tgggtcgtct cggcccgatc ggccaccgcg ctcgacgcgc agctcgcgcg actgcgggcg   46680
ttcgccgccg agcgcccgga cctgccgccc gccgacgtcg cccactcgct cgtcaccagc   46740
cgcgccacct tcgaacaccg ggcggtcctg ctggccgcgc ccgacggcat caccgcggcc   46800
gcccgcgccg aggcccgcga acgcagcacc gcgttcctct tctcgggggca gggtgctcag   46860
cggtcgggga tggggcgtga actgcatgct gctttcccgg tgttcgcggc ggcgttcgac   46920
gaggtggtgg cggtgttgga tgcggagttg gcgacgggtt ccggtggggg tgtgtcgctg   46980
cgggaggtga tgtggggcgg ggggtcggag ttgttggatc ggacgcgttt cacgcagccg   47040
gcgttgttcg cggtggaggt ggcgttgttc cgtttggtgg cctcgtgggg ggtggggcct   47100
gagttcgtgg cggggcattc ggtgggtgag attgcggcgg cgtatgtggc cggggtgttc   47160
tcgttggtgg atgcgtgtcg tttggtggtg gcgcgggctt cgttgatgga tgcgttgccg   47220
gtgggtggcg tgatggttgc ggtggaggcg gccgaggcgg aggtggtgcc gctgttggtc   47280
gatggggtgg cgatcgccgc ggtcaacggg ccggtttcgg tggtggtctc cggtgtggag   47340
gcggccgttg ggcaggtcgt ggatcagttg gtggagcggg gccggcgggt ccgtcggttg   47400
gcggtcagtc atgctttcca ctcgccgttg atggatccga tgttggatgc cttccgggcc   47460
gtcgccgagg gcctggagta ccaccagccg cgcatccccg tggtgtccaa cgtgacgggc   47520
gaggtggccg cggcggagga gctgtgcgcg gccgactact gggtgcggca cgtccgggcg   47580
accgtgcggt tcgccgacgg cgtccgcacc ctggccgagc gcggcgccac cgccttcctg   47640
gagatcggcc ccgacggcgt actgtccgcc ctggccgcgg cctgcctgtt cgacacggac   47700
gccgaagtgg tgcccgcgct gcgcaagggg cgccccgagg agcacaccgc cctcaccgcc   47760
gccgcccaac tccacgtggc cggcgtggac atcgactgga ccgcggtcct ggccggcacc   47820
ggcgggcggc ggatcgccct gcccacctat gccttccagc gcgagcggta ctggccctcg   47880
ctcgccgcac aggcccccgg cgacgccggc gggctcggcc tggaagccgg gcggcacccg   47940
ctgctcgggg ccgcgaccac cgtcgccgga tccgcggaga tcctgctcac cggccgcctg   48000
tccaccaccg cccagccgtg gctcgcggtc tacgaggcgg acggccgcac cgtcctgccg   48060
gccgcggtcc tcgccgaact cgccgtccgc gccggcgacc aggccgactg cccgaccgtg   48120
gcggaactga ccgtcgccgc accgctcgtc ctcaccggcg cggcggccca gcgcctccag   48180
gtccggggtgg ccgcccccga cgacaccggc cggcgcgcgc tgtccgtgca cgcccgaccc   48240
gacgactccc ccgacagccc ctggacgctg cacgccaccg cggtcctcac ccacgacacc   48300
ccgcagcccc cggcgccgga caccggctgg ccgccggagc gcgccgtgcc gctcgacgcc   48360
ctgcccaccg ccaccggccc ggcccggatc gcggcggcct ggcagtgggg cgacgaactc   48420
tgcgccgaga tcgaactccc cgaacccggc ccggcggagc gggcattcgc cctgcacccg   48480
gcgctgctgg acaccgcggt ccgcgccggc ggcctgctgg acggcgacgc caccctggac   48540
gccctcggct ggcggggcct cgccctgcac gccgcgtccg ccaccgccct gcgggtccgc   48600
ctcaccccgg acggcacgga cacctgggct ctggaggcca ccgacccgca gggcgctccg   48660
gtcgtctccg tcaccgggct caccctgggc acgcccaccg tcgaccggtc gggggccggg   48720
gcggccgatg acggcgcgac cctgctcgac ctggagtggg tgcccgcgcc gcaggccgcg   48780
cccaccggcg gcgaccacct cccgtacgcc gtgctcggcg atcaactcgc ggagctggac   48840
gggcagttga ggatcgcggg cgacgggccc gggcgcgtcg catcgctggc cgcgctgctg   48900
gacggcggtg cgccgctgcc ccggctcgtc ctcgcgccgg tgctgggcgt gccgaccggg   48960
gaaggcgacc tgcccgccgc ggtgcgcggc accaccacgg cggtgctgga gctgctgcag   49020
cgctggaccg ccgacgcccg caccgccgac agccacctgg tgatcgtcac ccgcggcgcc   49080
gtcgccgccg gggcggagga cgtgcacgac ctggcggcgg ccccggtctg gggcctggtc   49140
cgctcggcac agtccgaaca ccccggcagc ttcctgctgc tcgacctcga ccccgccgat   49200
cccgcgggag cctcccgcgc cgccgcgccg gccaccctgg cggccctgct cgacgcgggc   49260
gagacccagg ccgcggtgcg cgccgacacg ctcaccgtcg cccggctgac ccgggccgcc   49320
gacggacccg aggccaccgc cggacacccg gtgcgggact gggaccgcga cggcaccgtc   49380
ctgatcaccg gcggcaccgg cggcctgggc ggcctcctgg cccgccacct ggtcaccgga   49440
cacggcatca agcacctgct gctcgccggg cgccgcggcc cggacgcccc cggcgcgcgg   49500
gccctgcgcg acgaactggc cgccctcggc gccgaggtga ccgtcgccgc ctgcgacgtg   49560
gccgaccgtg ccgcactgga ccgactcctc gcgcaactgc cgccggagca cccgctgacc   49620
gccgtcgtgc acaccgccgg cgtcctcgac gacgccaccg tcggcaccct gacgcccgag   49680
cggctggaca ccgtcctgcg cgccaaggcg gacgccgcct ggcacctgca cgacgccacc   49740
cgcgaccgcg acctggcagg gttcgtgctg tactcctcgg tcgccggtgt caccggcggc   49800

```
cccggccagg gcaactacgc cgccggcaac acgttcctcg acgcgctcgc cgcgcaccgc    49860
gccgcccagg gcctgcccgg actgtcgctg gcctggggac cgtgggggca ggacgccggc    49920
atgaccggca ccctcggcgc cgccgacctg gcccgcctgg agcgctccgg catgccgccg    49980
ctcaccccgg aacagggcct ggccctgttc gacgccgccg gcgcccgcgg cgacgggttc    50040
gcggtggcgg tgcggctcgc ccgtggcgcc gccgcaccgg gcgccgacga ggtccccgcg    50100
gtgctgcgtg ccctggtgcg cggccggcgc cgcacggcgg ccgcggccgg gcacgccggt    50160
gtactggccc gccggctggc cgccctggac gccgagcagc ggcatcaggc gctgctcgac    50220
ctggtccgca ccgagacggc cgcggtgctc ggccactccg gggcggacgc cgtcccggcc    50280
gagcgggact tcaaccggct gggcttcgac tcgctgatgg cggtcgaact gcggacgcgg    50340
ctggccaccg ccaccggagc ccggctgccg gccacgctcg tcttcgacca cccgacgccg    50400
gacgcggtcg cccggcacct cgcgtcgacg ctgcccggtg ggaccgcggc cggtccggac    50460
cgttccccgc tggccgaact cgaccggatc gccgccgagt tgtcgccgga gggcgcggac    50520
gacgccaccc gacagggcgt cgtcgggcgg ctgcggcacc tgctggcgca gtgggacggc    50580
acccgacagg acggcggtgg gacgaccgtc gacgaccgca tcgaagcggc gagcgccgaa    50640
gaggtcctcg ccttcatcga ccacgagctc ggccggcagg cggactcctg acccgcccca    50700
ctcccgtcgc tcgcgcgcac cacatctgag gaaggtttca cggaccatgc cggacgaaaa    50760
gaagctcgtc gactatctga agtgggtcac gaaggacctc caccagaccc gccagcgcct    50820
tcaggaggtg gaggcggggc gccacgaacc cgtggcgatc gtcggcatgg cctgccgctt    50880
ccccggcggt gtgcgctccc cggaggacct gtgggagctg ctgtccgcgg gccgggacgg    50940
catcgggccg ttccccgccg accgcggctg ggacctggcg gcgctggccg gcgacgggcc    51000
cggtcgcagc gccacccagg aaggcgggtt cctgcccgac gcggccgcct tcgacccggg    51060
cttcttcgac atctccccgc gcgaggcgct cgccatggac ccgcagcagc ggctgctgct    51120
ggagaccgcc tgggaggccg tcgaacgctc cggcatcgac ccggccgggc tgcgcggcag    51180
ccgcaccggc gttttcgtcg gcaccaacgg ccaggactac gcgcacctgg tcctcgccgc    51240
gcaggacgac atgggcggct acgcgggcaa cggcctggcc gccagcgtgc tctccggccg    51300
actggccttc gcgctcggcc tggaaggccc ggccgtcacc ctcgacaccg cctgctcctc    51360
gtcactggtg accctgcacc tggccgcaca ggccgtgcgc gccggcgaat gcggcctcgc    51420
cctggccggt ggcgtcacgg tcatgacgac ctcgtcgagc ttcgccggct tcagcctcca    51480
gggcggcctg gcgccggacg gccgctgcaa ggcgttcgcc gaggcggccg acggcaccgg    51540
ctggtccgag ggcatcggcc tgcttctcgt cgagcggctc tccgacgcgc agcgcaacgg    51600
ccacccggtg ctcgccgtgc tgcgcggctc cgccgtcaac caggacggcg cgtccaacgg    51660
cctcagcgcg cccaacggtc cgtcccagca gcgggtcatc cgccaggcgc tggccggcgc    51720
cggactcgtc cccggcgacg tggacgcggt cgaggcgcac ggcaccggca cccggctcgg    51780
cgaccccatc gaggccggtg cgctgctcgc cacctacggc caggaccggc cgccgaccg    51840
gccgttgtgg ctcggctcgg tgaagtccaa cctcggccac acccaggccg ccgcgggcgt    51900
cgccggcgtc atcaagatgg tgctggccct gcggcatggc gtcctcccgc agaccctgca    51960
cgtggacgcg ccctcctcgc acgtcgactg ggagagcggc gcggtgcggc tgctcaccgc    52020
acccgtcgcc tggtccgagg gcgacgaccg ggtgcgccgg gccggcgtct cgtcgttcgg    52080
catcagcggc accaacgccc acgtcatcct cgaacaagcc cccgatcagc cggaaccgac    52140
cgcggaagag acggctgccg cggcgcccgg cggcaccgcc gaggagcggg ccgccgctcc    52200
cgtcgccccg cgcgccgtgc cgtggccggt cgcggcacgc accgccggcg ccctcgacgc    52260
ccaactggtc cgggtccgcg cgctgaccac cgcgcccggc cgcaccgccg ccgacgtcgg    52320
tcacgcgctg gccaccgccc gtaccccctt cgagcaccgg cgcgctgctgg tccacgaggg    52380
cggcgccgtc accgaggtgg cgcgcggcgc cgtccccacc ggtgaccggg cgggctggc    52440
cgtgctgttc tccggacagg gctcccaacg gccgggcatg gggcgcgaac tccacgcccg    52500
ctacccggtc ttcgccgccg ccttcgacga gaccgtcgcc ctgctcgacg cccggctcgg    52560
cacgtcgctg cgcgacatcg tctgggacca ggaccgcacc cggctcgacg acacccgcca    52620
cacccagccc gcgctgttcg ccgtcgaggt cgcgctgtac cgcctgctgg cctcctgggg    52680
catccggccc gaccacgtca ccggacactc catcggcgag atcaccgcgg cgcacgtcgc    52740
cggtgtgctg accctcgcgg acgcctgcac cctggtggcc gcccgcgcca ccgccatgag    52800
cgaactgccg cccggcggcg ccatggtggc gctggaggcc accgaggacg aggtgcgtcc    52860
gctgctcacc gacgacctcg cgatcgccgc ggtcaacgcc ccccggtccg tggtcgtcgc    52920
cggcgccgag gacgccgccc tcgccgtccg ccggcacttc gacgacctgg gccgccggac    52980
cacccggctc ccggtcagcc acgccttcca ctcgccgctc atggacccga tgctcgacgc    53040
cttccggacg gccctcgccc cgctgacctt cgccgagccg gagatcccgg tcgtctccaa    53100
cctcaccggc ctcccggcca ccgccgagga actcgccacc ccgcactact gggtgtgcca    53160
cgtccggcag gccgtccgct tcggcgacgg cgtgcgcgcc ctcgccgacc gcggcgtgcg    53220
gaccttcctc gaactcggcc cggacggcgt gctgtccgcc ctggtccggg agaacctccc    53280
cgagccgggc ctggtcgccg tgcccgtgct gcgcaaggag cggcccgagg agaccaccgt    53340
gctggccgcc ctgggaaccc tgtgggcgca cggcgcggac gtggactggg acgcggtgtt    53400
```

```
cgccggcacc cgcaccccgc aggccgaccc cgtcgagctg ccgacgtacg ccttccaacg   53460
cgcccgctac tggcccaccc tcggcgcccg ccacggcgac ccggccgacc tcgggcagac   53520
cgccgccgcc cacccgctgc tgggcgccgc cgtcaccctc gccgacgccg acgagaccgt   53580
gctcaccggc cgcctcgcgc tgccctccca cccctggctc ggcgaccacc gcagcgacgg   53640
ccggatcacc gtccccggcg tcgccttcgc cgaactcgcc gtccgcgccg gcgacctgag   53700
cggcaccccg cacctggcgc ggctcgacct gccggcgccg ctcaccctcg gcgacggcga   53760
caccgtcacc ctccaggtcc gggtcggcgc ccccgacccc gcgggccacc ggccgctgac   53820
cgtccacgcc cgcctcgcag ccaccgagga cgcccctgg accacctgcg cgaccggtct   53880
gctcgccccg gacgcccccg aagcgcccgc ggatccgatc ggcccggccg acgccgggtg   53940
gccgccgcgg gacgcccgcc cggtgcccgt cgccgacctc gacgcggccg ccaccgccgc   54000
aggccgccac tacggcccc atttccaggg cctgaccggg ctctggcggc gcgacggcga   54060
ggtcttcgcc gaggtggccc tgcccaccgc caccgccgcc gaccgcgcct tcggcatcca   54120
cccccgcgctg ctggccaccg cgctccgcgc caccgccgca ctggacgacg accacaccgc   54180
cggccacacc cccgaaccga ccggcatcac cggactcgcc ctgcacgcca ccggggccac   54240
cgcactgcgg gtccggctca ccgcgaccgg gcccgacacc gtggccctcg ccgccgcgga   54300
cgccacgggc ggcgcggtcc tgaccgccga caccgtcacc ctcggctccc cgcaggaccg   54360
cccggctccc gcaccggccg gccacaccgg gcagggcggc ctgttccacc tcgactgggt   54420
gccggtcgac cccggcagcc gagccaccgg cacccgctgg gccgtcgtcg gcgacgacga   54480
actcgacctc ggctacgccc tgcaccgcgc cgacgagacg gtcagtgcct acgcggcgtc   54540
gctgggcgga gccatcggcg acagcggtct ggcgcccgac gtcttcctcg tccccgtcgt   54600
cggcggcccg gacgccgggc ccgacgcggt gcacgccgtc accgcccgcg ccctggggct   54660
gctccaggag tggctgaacg agccgcggtt ggccggcgcc cgcctggtct tcgtcacccg   54720
cggcgccgtc gcggtgcccg gcgagaccgt caccgacccg gccggcgccg ccgtctgggg   54780
cctgctgcgc tccgcccaga ccgagaaccc gggcagtctg ctgctggtcg acctcgacga   54840
cgcgttccgg tccgccggga tgctgccgca cgtcctcacc ctcgacgaac agcagctcgt   54900
cgtccgcgac cacgcggtcc gcgccgcccg cctggcccgg ctgccggagc cggccgccgg   54960
caccgcgccg gcccgcgcct gggacccgga cggcaccgtc ctgatcaccg gcggcaccgg   55020
cggcctgggc gccgcgctcg cccgccacct ggtcaccgtc cgcggcgccc gccacctgct   55080
gctcgccggc cgccgcggcc ccgaggcgcc gggcgccggc gaactggtgg cggagctgac   55140
cgcacagggc gcggacgtgc gggtggccgc ctgcgacgtc ggcgaccgca ccgccctcga   55200
cgcgctcctc gccacggtcc ccgcggcgca cccgctgacc gccgtcgtgc acaccgccgg   55260
cgtcctggac gacgccctga tcggctcgct cacccccgac caactggcca ccgtgctacg   55320
gcccaaggcc gacgccgcct ggcatctgca cgacgccacc cgcggcctcg acctggccgg   55380
cttcgtcctg tactcctcgg tctccggcgt cctgggcagc cccggccagg gcaactacgc   55440
cgccgccaac gcctacctcg acgcgctcgc ccggcaccgc gccgaccagg gcctcccggc   55500
gctctccctc gcctggggcc cctggggtcg gggcagcggc atgaccgcgt cggtcagcga   55560
cgccgacctg gagcggatgg cgcgcggcgg cctgccgccg ctgaccgtcg aggacggcct   55620
ggccctgttc gacgccgccg tcggccgccc cgagccggcc ctggtgccca gccgcatcaa   55680
cgtcgccggc ctgcgggacc agcaggcact gccggcactc tggcgcgacc tggtaccgcg   55740
ggcccgccgc accgcggcca ccgccgaccg ctccccggtc acggtgcgcg agcgcctccg   55800
ccacctcgac gagaccggcc aggagcagct gctcatcgac ctcgtcgtcg gctacaccgc   55860
cggcctgctc ggccacccccg accccaccgc cgtcgatccc gaacggggct tcctggagct   55920
gggcttcgac tccctggtct cggtcggcct gcgcaaccag ctcgccgaga tcctcggcct   55980
gcgcctgccc tcgtccatcg tcttcgacag caagtcgccg gtgaagctgg cgcgttggct   56040
gcaccaggaa ctcgccaacg gcccccagcc gggcgccacc ggccccgccg ccgcggacgc   56100
ccgtcccgcg gtgcgctccg acgacaccct ggagggcctg ttctacaacg cggtgcgcgg   56160
cggcaagctc gtcgaggcga tgcggatgct caaggccgtc gccaacaccc ggccgatgtt   56220
cgacaccccc gccgagctgg aggagctctc cgagccggtg acgctcgccg acggcccggg   56280
ccggccccgg ctgatcttcg tcagcgcccc gggcgccacc ggcggcgtcc accagtacgc   56340
gcgcatcgcc gcgcacttcc gcggcagccg ccatgtctcc gcgctgcccc tgatgggctt   56400
cgcccccggc gagctcctcc cggccaccag cgaggccgcg gcccgtatcg tcgccgagag   56460
cgtcctgatg gccagcgagg gcgaaccgtt cgtcatggtc ggccactcca ccggcggctc   56520
gctggcctac ctcgccgccg gcgtcctgga ggacacctgg gacgtccggc ccgaagcggt   56580
ggtcctcctc gacaccgcgt ccatccgcta caaccccggc gagggcaacg acctggaccg   56640
caccacgagg ttctacctgg ccgacatcga ctcgccctcg gtgacgctca acagcgcccg   56700
gatgtccgcc atggcccact ggttcatggc gatgaccgac atccaggcgc cgcaccgac   56760
cgccccacc ctcctcgtgc gcgccgcccg ggccctcgac ggcttccggc tcgacacctc   56820
gtccgtcccc gccgacgagg tccgggacat cgacgccgac cacctctccc tcgccaagga   56880
gcactcggca ctgaccgcgc aggccatcga gggatggctc gcggaactgc cggaccccgc   56940
ggcctgatct ccggcccggc cggccccga caggccgggc ccatccccccc accggggcgg   57000
```

```
cgccacccag gtgcgccccg gcgcacccgg ccccgcgcag gcagccggag cgcccatccc   57060
cccgcatggc accaccctcc agaggagtcc ttccatgagc acaccgaccg caccgccctc   57120
cctgaaagcg gaggtgccgc ccgtcctgcg cctgagcccg ctgctgcgcg aactccagtc   57180
ccgcgccccc gtctgcaagg tccgcacccc cgccggcgac gagggctggc tggtgacccg   57240
gcacaccgaa ctcaagcagc tgctgcacga cgaccggctg gcccgcgccc acgccgaccc   57300
ggccaacgcc ccgcgttatg tgcacaaccc gttcctggac ctgctcgtcg tcgacgactt   57360
cgacctggcc cgcacgctgc acgccgagat gcgctccttg ttcacccccgc agttctcggc   57420
ccgccgcgtc atggacctga cgccgagggt ggaagccctc gccgaggggg tactggccca   57480
cttcgtcgcc cagggaccgc ccgccgacct gcacaacgac ttctcgctgc cgttctccct   57540
gtcggtgctg tgcgcgctca tcggcgtccc ggccgaggaa caggggaagc tgatcgccgc   57600
cctcaccaaa ctgggcgaac tcgacgaccc ggcacgcgtc caggaaggcc aggacgagct   57660
gttcggcctg ctgtccggcc tggcacgccg caagcgcatc acacccgagg acgacgtcat   57720
ctcccggctc tgcctgaagg tgccctccga cgagcgcatc ggcccgatcg cctccggtct   57780
gctcttcgcc ggcctggaca gcgtcgccag ccacatcgac ctgggcacgg tgctgttcat   57840
ccagcacccg gaccagctcg ccgcggccct ggccgacgag aagctgatgc gcggcgccgt   57900
cgaggagatc ctgcggtccg ccaaggccgg cggttcggtg ctcccgcggt acgcgaccgc   57960
cgatgtaccg atcggcgacg tgaccatcag ggccggcgac ctggtgctgc tggacttcac   58020
cctggtgaac ttcgaccgca cggtcttcga cgagccggag ctcttcgaca tccggcgcgc   58080
ccccaacccg cacctgacgt tcggccacgg catgtggcac tgcatcggcg cgccgctggc   58140
ccgggtcaat ctgcgcaccg cctacaccct gctgttcacc cgcctgcccg gcctgcggct   58200
ggtgcgcccg gtcgaggaac tgcgggtgct gtcggggcag ttgtcggccg gcctgacgga   58260
gctgcccgtc acctggtgac gtgatgtcgg acccgcccgg cccggttccg ggcaggcgga   58320
agtgagggcc cccgccccgc tcggccgggg gccctcacgc acggggagc gggctcctca    58380
gtccgccgcg acgctgatcg ccccggacgg gcagagcgcc gccgcgtcgc gtacgtcgcc   58440
gggatccgcg gcgtccgccg ccccggccag gacggtcacc agaccgtcgt cgtcctggtc   58500
gaacaggtcc ggtgcggtca ggacgcactg gccggccccg acacagcggc cggggtccac   58560
ggtgatgcgc acgatcgttc ctccgagggt cggttcgtcg ccggggtcgg gccgccgccg   58620
ggcggcccga ccgggctcac ttccaggtga cgggcagttc gtgcaggccg aacagcaccc   58680
cgtcgtactt caacggcagt tcctccaccg gcacggcgag cgtgagggag gggatccggg   58740
cgaagagcgt gcggtaggcg atgtccatct cctcccgcac caggttctgg cccaggcact   58800
ggtggacacc gtagccgaac gccacgtgcg agcgcgccga gcgctcgggg tcgaactccg   58860
agggcgccgc gaaggccgtg gcgtcgtggt tggcggcggc gatcagcggc acgatgccct   58920
cgccggcccg gatcagctga ccgccgatct ccacgtcctc gacggccacc cggaatgcca   58980
ccagatcggc caccgagtag tagcgcagca gttcctcgac gatccggtcg tcgccgatcc   59040
accgcgggtt ggtgagcaac tgcaccacgc ccaggccgat gttgttcgcc gtcgtctcgt   59100
gcccggcgat cagcagcagc atcgccaccc cggacagctc ctgccgggac agtgtgcccg   59160
cggcgatcag ccggctgatc aggtcgtcac cgggtcgccg ctgcttgatc tggatgagtc   59220
ggcccagata gcgcagcagc gcctgggtcg ccttgtcgcg ctcctcgtcg gtcgaactga   59280
gccggaccag cacccgggtc cggtcctcga agaagtcccg gtcgaccttg ggcaccccga   59340
gcagactgga gatcaccagg gacggcaccg gcagcgcgaa ggactcgacg aggtccgccg   59400
aggtgcccgc ggccagcatc gcgtcgatcc gctcgtccac gatctgctgg atcaccggcc   59460
gcagttcccg caccttgcgc acggtgaact cggggatcaa ggtcttgcgg aaccggccgt   59520
gctccggtgg gtccatcgcc acgaaccacc ccggcacctc gtactgcgac ggggccccac   59580
cggtccgacc ggccttgggg aaaccgggtt tggaggggtc ggcgctgatc cgggggtcgg   59640
tcagcaccgc ccggacgtcc tcgtgccggg tcaccaacca gaccgggccg ctggggagtt   59700
cggagcggac caggcccgcc ccgccgcggt aggtcgcgta ctccggcggg gggaagggcc   59760
ggcccggcct gcgcagcggg aaggccaccg ggcactgcgg cgcggcggtc gcgcgtcgg    59820
cttcggtgct catgccactc cgtagaacgc gcggatccgg ccggtgatga actcctgctc   59880
ctgtgcggtc agtccggtgt gcgtcggcag atagaagccg tcggcgctga gccgggccgc   59940
gttcaacgcg gccagtcgg cggagaagta accgggctga cggctcatcg gcttgaagaa    60000
gacccgggtc tcgatgccct cgccggcgag gtacgcgcac agctcgtccc gccgctcggc   60060
ccgcaggtcg tacatccaca gcacgtcgcg cggcggcatc agcgtgatgc cggggatgtc   60120
gcgcagcgcc tcgtcgtagc gcttctcgat gtcgcggcgc agggcgagga tggtgtccag   60180
ctgctcggtc tgcgcgagcg ccaccgcggc ctgcatgttg gtcatccgga agttgtaggc   60240
cagcttcttg tgcaggaagc tgtggtcctt ggtgaacgcc atcgcccgca gatgggccat   60300
ctgctcggcc aggtgcgggt cgtgggtcag gcagacgccg ccctcgccgg ccgagatgat   60360
cttgttggcg aagagcgaga aacaggcgat gtcgccccgc ggccgcaccc cgtgcgcctc   60420
ggcggagtcc tccaccaccc gcaggttgta ctcgtacgcc aggttcagca cggcgtccat   60480
gtcgcactgc cggccgtaga tgtgcaccgg catgatcact ttggtgcgcg gggtgatctt   60540
ctcctcgatg cgggacacgt cgatgttcag gtcgtcgccg cagtccacga acaccggcgt   60600
```

```
ggcaccggtg taggtcaccg cccaggcgga cgcgatcatc gtgaactccg ggacgatcac   60660
ctcgtcaccg gggccgacgc ccagcgcgcg cagcgccagc gtcagcgccg tggtgccgga   60720
ggagcaggcg acgccgaacg gcacgtcgtt gtacgcggcg aacgcctcct cgaaccgcct   60780
gacgtacggc ccctgcgaag agatccagcc gccgccgacg gcctccgtca catagtcgag   60840
ctcgcggccc tggagccacg gcatggacac cggatacgta aaggacatgg gttttgagtc   60900
ttcctcggtc agtcggttgc cagggcggga aggccgagca gcaggtccgc cgccgcggcc   60960
cggccgccgg cgtcccgcag cagcccggcg aagtgctccg cgcgctcggt gaaggagggt   61020
tggtcgagca cgcgggtgat cttgtccagg acgtcctcgg tgtccacggt ctccggccgg   61080
tccagggtca ggctcacccc gaagtcctgg ccccggatcg cctggtcgtc gcagtccacc   61140
cacaacggcc ggaccaccag cggctttccg aagtacaggc cctcgtggta gccgttgcca   61200
ccggcatggg tgaagaacgc cttcacgttc ggatgggcca gcacgtccag ctgcgacggc   61260
acccagccct cgatccgcag gttgtccggc agctcggcgg ccggcggcag caactcctgt   61320
tggccgcgcg ggagtttcca caacacctgg tggccccggc cgtccagtcg ccgggcgacc   61380
tccaccagcg acgccacctg ctcacgggtc agccgggtga tcgtgccgaa gcccatgtac   61440
accacggact tctgcgccga cagccagtcc gacaggccgt cgtcgtccgg tgcctggggc   61500
agcggcggca ccatcgtgcc caccagccgc agcttcggat gcatcgggaa cgggtagtcc   61560
aactccctta cggagtagca caagacctgc tccgcatggt cgatccgcgc catcatctgc   61620
cgcgcctggg cgcgcgatgcc cagctcggtg cggacccggt tgtcctcctc gacgaccttg   61680
cggacgtccg acgtcaggaa catcccgagc gtccgcagcc ggaacagctg gttctcgatc   61740
cgctgagcca gggacatcgc ggccggcagc cccgagtgcg gcaccgggaa acccgacggg   61800
gtgtaggact tggcgaacgg gacgtgcgag gtgaggacgt tgctcggcac gaacggcacc   61860
ccgagcacga acggaatgcc cttggtgatc gccagctcgt acccgaactg gcacatgctc   61920
tcgatcacca tcagcgccgg ctcgacctcc tcgacgatct cctccaggcg gcggtacttc   61980
gccatccgcg actccggcgc gaacgaatgc cgaatcaccg ccgcgtgcgc cttgaaccgc   62040
gaccgctgcg tcacctccgc atacgtcgcg tcgtcccatg tgaccgccga catctgcgag   62100
acggtgtcgc cgagcgacgc gaaccgaacc gggctgccgt ccaccacggc cgccacctcg   62160
tcgcgcgctt tctcgtcggt ggcgaaccac aggtccgcca cgtcgcgccg ggacaattcc   62220
ccggccagca cgagcagcgg attgagcagg ccgctttcgg cataactgac gaacaggatc   62280
ggccgccgat tcgcgcccat ggacaacacc cctcggaatg tggcgggccg ccgggcccgc   62340
gcgccacgca cccgcccggc ccggtcgccg ggtgagtgca ttcgccgacg ccgccacccg   62400
aggcgcgtgt tgccggaagg aagggtcacc ggccggcacc cggaacgcgc cgcgtggaaa   62460
acgggtcggt tacttggtct catgccacgg accggggaat cactagtctt cggcgcgcga   62520
cggccctttc cgggccgtgt ggccaatgcc cgtccccggc gcccgtcatt ccttagggaa   62580
aagtacagcg tttgcgaacg tacgatccgg cacgcagagg tgacctgagg ccaacttttc   62640
cgcaggggtg agcaaggcat gacgatcgga gccgacgagg acccggtggt ggtcgtcgga   62700
atggcctgcc gttatccggg tggggtcgcc ggcccggagg acctgtggga actggtccgc   62760
accggccgcg acgcgaccac cgccttcccg gacgaccgcg gctgggacct ggccgcactg   62820
gccggcgacg gacccggccg cagcgcgacc cgcgagggcg gattcctcac cggcgccgcc   62880
gacttcgacg ccgccttctt cggcatgtcg ccccgcgagg ccgtctccac cgacccgcaa   62940
cagcgcctcg tcctggagac cgcctgggaa gccctggagc gcgccggcat cgacccgcac   63000
tccctgcgcg gcagccgcac cggggtcttc gtcggcgcca gcggccagga ctacgccgcc   63060
gtcacccacg cctcgcccga cgacctggac ggacacgccc tcaccggcct ggcccccggc   63120
gtcgcctccg gtcgcctggc gtacgtcctg ggcctcgaag gccccgccgt caccgtcgac   63180
accacgtcct cctcgtcgct ggtcgcgctg cactgggcgg tccgcgccct gcgcgcgggg   63240
gagtgcagca ccgccctggc cggcggcgtc acggtgatgt ccaccccggc cgccttcgtc   63300
ggccacaccc gacagggcgg cctcgcgccc gacggccgct gcaagccgtt ctccgacgac   63360
gccgacggca ccgcctgggc ggagggcgtc ggcatcgtcg tcctggagca cctgtccacc   63420
gcccgcgccg ccggcaaccc cgtcctcgcc gtgctgcgcg gctcggccgt caaccaggac   63480
ggcgcctccg acggcctcac cgcacccagc ggtcccgccc aggaacgcgt catccgcgcc   63540
gccctcgccg acgcccgact cgcccccgcc gacatcgatc tcgtcgaggc gcacggcacc   63600
ggcacccggc tcggcgaccc cgtcgaggcc cgggcgctgc tcgccgccta cggccaggac   63660
cgggacccgg accgaccgct gcgcctcggt tccctgaagt ccaccctcgg ccacgcacag   63720
gccgccgccg gcatcggcgg agtgatcaag accgtcctga ccctgcggca cggcctgatg   63780
ccgcgcatcc ggcacctggc cacccccacc gccaagtcga ctggtcccag gggcgccgtg   63840
gcccccctca ccgaccacac gccctggcca ccggccgacc gaccgcgccg cgccggcgtc   63900
tcctccttcg gcatcagcgg caccaacgcc catgtgatcc tcgaagaggc gccgcccgcc   63960
gacgtccccg tcacccggcc cggcaccctc cgccccagca ccgtcccctg gccggtctcc   64020
gccgccacgc ccgaagccct cgacgcccaa ctcgcccggc tccgcgccca cctgcgcacc   64080
cactcggacc tggacccgct ggacgtcggc tactccctgg ccaccggccg cgccgcgctc   64140
cgccaccggg cggtcctcct gccgcccgcc gacggcaccg ccgcggacgc cgtcgagcac   64200
```

```
gcccgcggtg cggcccacca gcgccgcacc gccgtcctct tctccggcca gggcagccag   64260
cgcccgggca tgggccgcga actcgccgcc cgcttcccgg tgttcgccga cgcactggac   64320
gacgcgctgc gcgccctgga ccggcacctg gacggcccgg tgcgcgaggt gatgtggggc   64380
accgacgccg cgctcctgga ccggaccggc tggacccagc ccgccctgtt cgccgtcgag   64440
gtcgccctcc accgcctggt cgcgtccctc ggcgtcaccc ccgacttcgt cggcggccac   64500
tccgtcggcg agatcgccgc cgcccacgtc gccggcgtcc tgagcctgga ggacgcctgc   64560
cgcctggtgg ccgcccgcgc cacgctgatg caggcgctcc cggccggcgg cgcgatggcc   64620
gcgctggagg ccaccgagga cgaagtggcc ccgctgctcg gcgcacacct cgcgctggcc   64680
gccgtcaacg gccccaccgc ggtcgtcgtc gccggagccg aggacgccgt gcggcaactg   64740
accgcccgct cgccgaccg cggccggcgc accagccggc tggccgtctc gcacgccttc   64800
cactcgccgc tgatggagcc catgctcgac gccttccggg acgtcgtgag ccgactgacc   64860
ttccaccagc cgtcgatccc gctggtctcc aacctcaccg gtgaactcgc cggcagtgag   64920
atcaccagcg ccgagtactg ggtccggcac gtccgcgaca ccgtccgctt cgccgacggc   64980
atcaccgcac tggccaaggc cggcgccgac gtcctgatcg aactcggccc cggcggcgtg   65040
ctgtccgcga tggcccgcga caccctcggc cccgacagca ccaccgacgt cgtccccgcc   65100
ctgagcaagg gacggcccga ggagaccgcc ttcgccggcg ccctcggccg cctgcacacc   65160
ctcggcgtcc ccgtcgactg gcccgccttc tacgccggca ccggcgcccg ccgcgtcgaa   65220
ctgcccacct acgccttcca gcacgtgcgc cactggccca ccccgccccg cccgaacggc   65280
gccgggcccg gcgccctcgg ccaccccctg ctcggctccg ccgtcgaact cgccgacggc   65340
ggcggcaccg tctgctccgg cgccctctcc ctccgcaccc accctggct cgccgaccac   65400
accgtcgccg ggcgggtcgt gctgccggcc accgcgctgc tggaactcgc cgtgcgcgcc   65460
ggcgacgagg cgggctgcga cgtcctgcac gaactccacc tcaccacccc gccggccctg   65520
cccgacgacg ccgccctgca cgtccaggtg cacgtcggcc ccgccgacac caccgggcgc   65580
cgcgccgtca ccgtccacac ccgcccccgac caccacccgg ccggcgactg acccgatgc   65640
gccaccggca ccctcggcag caccccgccg tccgcagccg aagccgccac gggcggcacc   65700
ccggccgcct ggccgccggc cgacgccgaa cccctcgacc tcgccgacca ctacgaacgg   65760
ctcgccgacc gcggcttcga ctacggcccg accttccgcg gcctgcgggc cgcctggcga   65820
cgcggcgcgg agatcttcgc ggacgtggaa tgcccgcccg gcaccgccga cgacgccccc   65880
gaccacggac tgcaccccgc cctgctcgac gcggcccggc acgccgccat ggcggtggac   65940
ggcaccgtgc ccgtcgcctg gcacggcgtc cggctgcacg ccgtcggcgc caccgcgctg   66000
cgggtccgca tccgccccac cacgaccggc acgctgaccc tcaccgcggt cgacgtgcac   66060
ggcgcgccgg tcgtcaccgt cgaggccctc accgcccgcc cgctgaccga cgaggaacgc   66120
gccgccccgc ggacgccgcg gcaggcccgc ggcgagacgc ccgccgacgc ccgcccggcc   66180
cggcccgcgg cggcccgccc cggcccggcc ggcgaacccc tcccggacac caccgggtcc   66240
cacccaccg ccggccacct cgccgcgctg ccgccggccg cccgggagcg ccagctgctg   66300
gacctggtgc gcacccaggc cgccgccgtc ctgggccacc ccggccccga ggccgtcggc   66360
acccgcagcg tcttcaagga gctgggcttc gactcgttgg ccggcgtcga actcgccgac   66420
cggctcaccg cccgcaccgg actgcgcctg ccggccaccc tcgtcttcaa cttccccacc   66480
cccgaacgtg ctgcccaccg cctcggggaa ctcctcgccg caaccgcccc cctcgacccc   66540
ggggcgtacg gagaggaact caccaggttc gaggcgatcg tgacgaacct gccgcaggac   66600
ggccccgaac gccgggccgt cgcggaccgg ttggacgcca tcgtctccgc actccgccag   66660
aactcgcctg cagaggtgcc ctcctcggac gaggacatcg acacggtgtc ggtcgacaga   66720
ctgctcgaca tcatcgatga agagttcgaa accacataga gaaattgttg ctttcgttcg   66780
cgacccgatg acgaggacgg accgatgcag gaaccccagc aaggccagcc ggaccagcag   66840
gagaaaatcg tcgactatct ccggcgggtc acttcagatc ttcgccgtgc ccgccgccgc   66900
attggcgaac tggaatccaa ggacaacgag cccatcgcca tcgtcggaat gggctgccga   66960
cttcccggcg gcgtcaattc gccggaatcc ctgtgggacc tggtgcgttc cggcggcgac   67020
gccatttccg gattccccgt cgaccgcggc tgggacctgg aaaccctcac cggaaacggc   67080
gacggcagca gcgccaccca cgaaggcgga ttcctctacg acgccgcgga attcgacgcc   67140
gccttcttcg gcatctcgcc gcgcgaggcg actgctatgg accccagca gcgcctcctc   67200
ctcgaagtcg cctgggaggc gctggagcgc gccggcatcg cccccacagc cctgcgcgg   67260
agccggtccg gcgtgttcgt cggctcctac cactggggcg cgccctcggc cgacgccgcc   67320
accgaactgc acggccacgc cctgaccggc accgccgcca gcgtgctgtc cggccgcctg   67380
gcctacaccc tcggcctcga aggcccggcc gtcaccgtcg acaccgcctg ctcctcctcc   67440
ctcgtcgccc tgcacctggc ggcccagtcc ctgcgcgtcg gcgaatcctc gctcgccgtg   67500
atcggcggcg tcacgatcct caccgagccg tccgtcttcg tcgagttcag cgcccagggc   67560
ggcctggcac cggacggccg ctgcaaggcg ttctccgacg ccgccgacgg caccggttgg   67620
gccgagggcg tcggtgtcct cgtcgccgag cggctctccg acgcgcagcg caacggccat   67680
ccggtcctcg ccgtgctgcg cggctccgcc gtcaaccagg acggcgcctc caacggcctg   67740
accgcccccca acggcccctc ccaggaacgg gtcatccagc aggccctcgc ccggaccggc   67800
```

```
ctgacccccg ccgacatcga cgccgtcgag gcgcacggca ccggcacccg gctcggcgac   67860
cccatcgagg cccaggccct gctcgccacc tacgccagg gacacacccc cgaccagccg     67920
ctgtggctcg gctccctgaa gtccaacatc gggcacaccc aggcggccgc cggcgtcgcc     67980
ggtgtcatca agatggtcat ggcgctgcgc cacggccacc tgccgccgac cctgcacgcc     68040
gacgcgccct cctcgcacgt ggactggtcc gccggatcgg tacgcctgct gaccgagggc     68100
cagcagtggc cggagaccgg acgtccgcgc cgggccgcgg tgtcctcgtt cggcatcagc     68160
ggcaccaacg cgcacgccct gctggaacag gcaccccacc ccgcggacac cgcggacgcc     68220
ggcgacgacg ccgcgcccac cgaaccggcc ggcgcgcccg ccgcgctgcc ctggatcgtc     68280
tccggacact ccccgcaggc gctgcgcgac caggccgccg ccctggccgc cagggtcgag     68340
accgaccccg cgctccgccc ccaggacatc gggcacaccc tgcacaccgc ccgcgccctg     68400
ctcgaacgac gcgccgtcgt cgtcgccccc gaccgcgccg aactcctcgc ggctacccac     68460
gagttggccg ccggccggtc cgcgaacgcc gtcgtcgagg gcctcgcgga cgtcgagggt     68520
cggacggtgt tcgtgttccc cggtcagggt tcgcagtggg tggggatggg ggcccaactc     68580
ctcgatgagt cggcggtgtt cgcggagcgg attgccgagt gtgcggcggc actcgccgag     68640
ttcaccgact ggtcgctggt cgatgtgctg cggggtgtgg tgggtgcgcc gtcgttggag     68700
cgggtcgatg tggtgcagcc ggcgtcgttc gcggtgatgg tgtcgttggc tgcgttgtgg     68760
ggttcccgtg gtgtgttgcc ggatgcggtg gtggggcatt cgcagggtga gatcgctgcc     68820
gcggtggtgt cgggtgcgct gtcgttgcgg gacggggcgc gggtggtggc gctgcggagt     68880
caggccattg gtcgtgcgtt ggcggggcgg ggcgggatga tgtccgtcgc gctgtcggtg     68940
gacgtgctcg aaccgcggtt ggtcgagttc gaggggcggg tgtcggtggc cgccgtcaac     69000
ggcccgcgct ccgtcgtggt cgccggcgag cccgaggcgc tggacgcgct gcacgcccgg     69060
ctgaccgccg acgacatccg ggcccgccgg atcgcggtgg actacgcctc gcactcgcac     69120
caggtcgagg acctgcacga ggaactgctg gaggtgctgg cggagctggc gccgcgcacg     69180
tcggaggtgc cgttcttctc gaccgtgacc ggcgactggc tggacaccgc gcggatggac     69240
gccggctact ggttccgcaa cctgcgcgga cgggtgcggt tcgcggacgc ggtggcggac     69300
ctgctggcgg cggagtaccg cgcgttcgtc gaggtcagct cgcacccggt gctgacgatg     69360
gcggtcttgg acctgatcga ggaggccggg gtcacggccg tcgcgaccgg caccctgcgc     69420
cgtgaccagg gtggcgcggg ccgcttcctg ctgtcggccg ccgaggtctt cgtgcgcggt     69480
gtggacgtgg actgggcggg ggcgttcgag gggaccggtg cggcccgggt cgacctgccc     69540
acctacgcct tccagcgcga gcggtactgg aacacccgca ccgccgccga ccgcacccg     69600
gccgacgccc cgatggacgc cgaattctgg gccgccgtcg aacaggcgga cgtctccgcg     69660
ctgaccgccg cgctcggcac cgacgaggac tccgtcgccg ccatcctgcc cggcctcacc     69720
tcctggcgcc gggcccgctc ccagcgcacc accctcgact cctggcgcta ccgcgtcacc     69780
tggacgcccc tcgcccaggt gccccgcgcc accctgaccg gcacctggct gctggtcacc     69840
accgacggca tcgacgacac cgatgtggca ggggcgttgg agagctacgg cgccgaggtg     69900
cgccggctgg tcctggacga ggagtgcacc gaccgcgccg tcctgcggga gcggctggcc     69960
ggcgcggagg acgtgaccgg catcgtctcc gtcctcgccg ccgccgagga cgacgccgca     70020
cgccaccccg gcctcacccg gggactcgcg ctcaccgtct ccctcgtcca ggccctgggc     70080
gacgccgagg cgaccgcgcc gctgtggttc ctgacccgcg cgcgccttcgc caccggcccg     70140
tccgaccccg tcacccggcc cctgcagagc cagatcgcgg cgtcggctg gaccaccgcg     70200
ctggagcacc cgcagcgctg gggcggcacc gtggacctgc ccgacaccct cgacgcccgg     70260
gccgcccagc ggctcgccgc cgcgctgtcc ggcgccctcg cgccgagga ccagctcgcc     70320
gtccgcgccg ccggggtact ggcccgccgc atcgtgcgtg ccggacaccg cgccggacga     70380
ccggcacgga cctgggcgcc gcgcggcacc accctgatca ccggcggctc cggcacctc     70440
gccccgcagc tcgcccgctg gctggccgaa cgcggcgccg agcacgtggt gctggtcagc     70500
cggcgcggtg ccgacgcccc cggagcgccc gaactcatcg cggaggcagc cgagtcgggc     70560
accgaggtga ccgtcgccgc ctgcgacatc accgaccgcg acgcggtcgc cgcgctgctg     70620
gccgacctca cggccgacgg ccgcaccctg cgcaccgtca tccacgccgc cgccgccatc     70680
gagctgtccg cgctcgccga caccaccgtg gcggagttcg ccgacgtcgt gcacgccaag     70740
gtcaccggcg cacggatcct cgacgaactg ctcgacgacg cggaactgga cgacttcgtc     70800
ctgtactcct ccaccgccgg catgtggggc agcggcgtgc acgccgccta cgtcgccggc     70860
aacgcctatc tgtccgcgct cgccgagcag cgccgcgccc gcggactgcg caccacctcc     70920
atccactggg gcaagtggcc cgacgaccgg gcacgcgagc tggccgaccc gcaccggatc     70980
cgccgcagcg gtctggagta cctcgacccc gagctggcgc tcaccgcgct ccagcacgtc     71040
ctggacgacg acgagaccgt catcggcctc atggacatcg actgggacac ctaccacgac     71100
gtgttcaccg cgggccggcc cgcgcacctc ttcgaccaga tccccgaggt gcggcgccgc     71160
ctcgaccagg catccgtccc ggaccccgcg ggcccggccg ccgacggcct ggccgcccgc     71220
ctgcacggcc tcgccgccgc cgaacaggac cggctgctgc tcaccctggt ccgcaccgag     71280
gccgccgccg tcctcggcca cgcctcggcc gagtccttcc ccgagcgccg cgccttccgt     71340
gacctcggct tcgactcggt caccgccgtg gacctgcgca accggctcgt ggccggcacc     71400
```

```
ggactgcggc tgccctcgac gatggtcttc gaccacccca actgcgcggc gctcgccgcg   71460
ttcctgaaga cgacggcgct cggcgtcccc ggcgccgcac cgcagcagca cgccgctacc   71520
ggcaccccgg ccgacgacga cccgatcgcc gtgatcggca tgagctgccg ctaccccggc   71580
ggcgccgcca cccccgagga actgctgcgg ctcgccctcg acggcgccga cgtcatctcg   71640
gagttccccg cggaccgcgg ctgggacgcc cggggcctgt acgacccgga ccccgaccgc   71700
cccggccaca cctactccgt ccagggcggc ttcctccacg aggccgccgg cttcgatccc   71760
ggcttcttcg ggatctcccc gcgcgaggcg gtcgccatgg acccgcagca gcggctcctg   71820
ctggagacct cctgggaggc gttcgaacgc gccggtatcg accccgcgtc actgcgcggc   71880
agcgccgccg gcaccttctt cggcgccagc taccaggact actcctccac cgtgcagaac   71940
ggcacggggg agtccgaggc gcacatggtg accggcaccg cggccagtgt cctgtccggc   72000
cggggtctcct acctgctcgg cctggagggc cccgcggtca ccgtggacac cgcctgctcc   72060
tcctcactgg tcgccctgca cctggcctgc cagtccctgc gcgacggcga gagctccctc   72120
gcgctggccg gcggtgcggc cgtgatggcc accccgcacg cgttcgtcgg cttcagccgg   72180
cagcgtgccc tggccaagga cggccgctgc aagccgttct ccgacaccgc cgacggcatg   72240
acgctcgccg agggcgtcgg cgtcgtcctg ctggagcgcc tgtcccacgc ccgcgccaac   72300
gggcaccggg tactggccgt gatccgcggt tccgccgtca accaggacgg cgcctccaac   72360
ggcctgaccg cgcccaacgg cccgtcccag cagcgcgtca tccgccaggc gctcgccaac   72420
gccggcctga ccggcgccga cgtcgacgcg gtcgaggcgc acggcaccgg caccaagctg   72480
ggcgacccca tcgaggccca ggccctgctc gccacctacg gccaggaccg cgacgccgaa   72540
cggccgctgc tgctgggctc ggtgaagtcc aacatcggcc acacccaggc cgcggccggc   72600
gtcgccggcg tcatcaagat ggtgctggcc atggacgccg gcgaactgcc cggcaccctg   72660
cacctcgacg cgccctccag ccacgtcgac tggaccgccg gcgccgtgga actgctgcgc   72720
gggcgcaccc cgtggcccga gagcgggcgc ccccgccggg ccggtgtctc ctcgttcggc   72780
atcagcggca ccaacgccca cctgatcctc gaacaggccc cggccaccga ccgccagcc   72840
gaccccgacc gcctccggga caccgccacc gacaccgtcg tcccctggcc gctcgccgcc   72900
aagtccccgg ccgccctgcg cgcccaggcc gcccggctcc tcgccaccgt cgagcacgac   72960
cccgacctcc cgcccgcccc cgtgggccac gccctggcca ccacccgcgc cgccctcgaa   73020
caccgcgccg tcgtcgtcgg cgagcgccgc gaggacttcc tgcgcggcct ggccgccctg   73080
tccaccggcg cctcgacggc cggcctggtc agcggcatcg ccggccccga ccccgaggga   73140
gcggtcttcg tcttccccgg ccaggatcc cagtggtggg gaatgggccg cgaactcctc   73200
gccacgtccg aggtgttccg caccgcgatc gatgactgcg cgacggccct cgccccgtac   73260
gtcgactggt cgctgcacga cgtcctggcc ggcgagggcg accccgccct gctggagcgg   73320
gtggacgtgg tccagcccgc gctgttcgcc atgatggtcg ggctgtccgc gctctggcgc   73380
tcccacggcg tcgtcccggc ggccgtggtc ggccactcgc agggcgagat cgccgcggcc   73440
tgcgtcgccg gagccctcag cctggccgac gccgcccgcg tggtggcgct gcgcagccag   73500
gcactgccgc aactgtccgg acgcggcggc atgatgtcgg tctccgcccc cgtagagcgg   73560
gtcaccgcac tcctcgcccc gtggcaggag gcgctgtccg tcgccgcggt caacggcccc   73620
tcgtccgtgg tcgtctccgg cgacaccgac gcgctcgacg ccctgcacac cgcctgccag   73680
gaacagggcg tgcgggcccg caaggtgtcc gtggactacg cctcgcacgg cggcacgtc   73740
gaggccgtcc gcgacgaact cgcccgcgtc ctcgcgccgg tcgacccgcg cgcccccgag   73800
gtgccgttct actcgacggt caccggcgac cgcgtggacg acgccgcctt cgacggcgcc   73860
tactggtaca ccaacctccg ccagaccgtc cgcatggagg aggccacccg cgccctcctc   73920
gccgccggac accgcgtctt catcgaggtc agcccgcacc cggtgctcgc cgccccgatc   73980
caggagacgc aggaggccgt agcggaggcc accggcgggt ccgcggtggt cctcggctcg   74040
ctccgccgcg acgagggcgg cccgcggcgc ttcctgacgt cgctcgccga ggcccacacc   74100
cacggcgccc cggtcgactg gaccaccacc ttcgcccggt ccgcctacca gccggtggac   74160
ctgccgacct accccttcca acgacaggac ttctggcccg aggcccggcc cgccaccccg   74220
gccgccggcg ccgacgcgtc cgacgccgcg ttctggcaac tggtcgagaa ccaggacctc   74280
gccgcgctcg ccgacgcgct cggcgtcccc gccgacgacg agcacaccgc gctcggcacc   74340
gtgctgccgg ccctgtccgc ctggcgcgcc aaggcccagg cccgcacccg gatcgacgaa   74400
ctccgctacc acgtccagtg gacccgggtc gccgagcccg cggcggcccc caccaccggc   74460
cggctgctgg tcgccgtccc gccggaccac gccgacgccc cctgggtcgc cgcggcgctc   74520
gacgccctgg gcaccgacac cgtccgcttc gaggccaagg gcaccgaccg cgcgggatgg   74580
gccgcacaga tcgcccaact cgtcgaggac ggcgaggagt tcaccggcgt ggtgtcgctg   74640
ctggccgccg ccgaggatct ccacccggac ttcggctcgg taccgctggg gctggggcag   74700
accctcgtcc tcgtccaggc cctcggcgac gccggcctga ccgcgcccct gtggtgcctg   74760
acccgcggcg ccgtcgccac cggccgcgac gacgccctcg acagcccgac ccagggcgcc   74820
ctgtggggcc tcggccgggt cgtggccctg gaacacccg accgctgggg cggcctgatc   74880
gacctgcccg ccaccctcga cgcccgcgcc gcggcccgcc tcaccggcct gctcgccgac   74940
cccgccggtg aggaccaact cgccgtccgc gccaccggcg tgctcgcccg ccgcatggtg   75000
```

```
cacgccgcgc cgtccgcgcc ccgcaccggg cgccgctggc ggggccgcgg cacctgcctg    75060
atcaccggcg gcaccggcgg catcggcggc cgggtcgccc gctggatggc cgagcacggc    75120
gccgcccacc tggtcctgac cagccggcgc ggcccggacg caccgggcgc cgccgcactc    75180
cgggccgaac tggaggccct gggcgcccgc gtcaccctcg ccgcctgcga cgtcgccgac    75240
cgcgacgccc tggccgccct gctggccgac ctccccgccg accagccgct cacctccgtc    75300
ttccactccg ccggcgtggc cgacggggac gcccgggcag ccgacctgac cctagatcag    75360
ctcgacgcgc tgctgcgcgc caaactgacc gccgcccacc acctgcacga gctgaccgcc    75420
cccctcgacc tcgacgcgtt cgtgctcttc tcctccggcg ccgcggtctg gggcagcggc    75480
ggccagcccg gctacgccgc cgccaacgcc tacctcgacg ccctcgccgc ccaccgcagg    75540
tccctcgacc tgcccggcgc gtccgtcgcc tggggcacct ggggcgaggt cggcatggcc    75600
accgtccccg aggtccacga gcgactgcac cgccaagggg tccgcgccat ggaaccggac    75660
cacgcgatcg gcgcgctcca gcagatgctg gaggacgacg acaccaccct cgccgtgacc    75720
ctcatggact gggaggcgtt cgcaccgagc ttcaccgcga cccgacccag cgccctgttc    75780
agcacggtgc ccgaagccgt ccgcgcggtg accggcgacc cgggcaccac ggccggcgac    75840
gacgtggact ccgcgacccc gccgctccgc cgccacctgg aggagctgtc cgccgccgag    75900
cgcggccggg ccctggtcga ggcggtccgc gccgaggcgt ccgcgaccct cggccacgac    75960
acccccgacg ccatccccgc cggccgtgcc ttccgcgacg tcggcttcga ctcggtcacc    76020
gccgtcgaac tgcgcaaccg gctgcgcacc gccctcggcc tgccgctgcc ggccgcgctc    76080
gtcttcgacc accccacccc cacggccctc gccggccacc tcggcgcgct gctcttcggc    76140
accgcccccg aggacgccgg caccggccgc cccgacgacc ccgacgcccg catccgcgag    76200
gcgctcgcca ccgtccccat cggacggctg cgcaaggcgg gcctcctcga catggtgctg    76260
aaactcgccg acggagacgc gaccgacgcc cccgcccccg aggccgacgc cccctcggaa    76320
tccctcgacg acatggacgc cgaagccctg ctgcggctgg ccaccgagaa ctcggcgaac    76380
tgaaagagag ctggagccca ccatgagcac gaaccccgac aagtacgtcg aggcactccg    76440
gtcgtccctg aaggagatcg agcggctgcg ccggcagaac gaacagctgg tggcggcggc    76500
ggtcgagccc gtcgcggtcg tcggcatcgg ctgccgcttc cccggcggcg tcacctcccc    76560
cgaggacctg tgggagctgg tcgccgaggg gcgcgacgtc atcgggccgt tcccgcagga    76620
ccgcggctgg gacctggaga agttggccgg cggcggcgag ggcggcagcc tcgcgcaggt    76680
cggcggcttc gtcgaggacg ccgccggctt cgaccccggc ttcttcggca tctccccgcg    76740
cgaagcggtc gccatggacc cgcagcagcg catcctgctg gagatcacct gggaagccct    76800
ggaacgcgcc ggcatcgacc cgtccaccct gcgcggtacc cccaccggcg tcttcgtcgg    76860
caccaccggc caggactacg gcgaggtcat caaggcgtcc gccgaggacg tcgaggtcta    76920
ctcgaccacc ggccacgccg ccagcgtcat ctccggccgg ctctcctaca ccctcggcgc    76980
cgagggcccg gccgtcaccg tcgacaccgg ctgctcctcg tccctggtcg ccctgcactg    77040
ggccgtccag gcgctgcgcg gcggcgagtg ctccatggcc ctggccggcg gcgcgtccat    77100
catggccacc ccgggcccgt tcgtcgcctt caccgcgcag agcggcctgg ccgccgacgg    77160
ccgctgcaag cccttctccg accgggccga cggcaccggc tggggcgagg cgccggcat    77220
gctggtcctg atgcggctct ccgacgccca gcgcgagggc cgcccggtcc tcgccgtgct    77280
gcgcggctcc gccatcaacc aggacggcgc ctccaacggc ctgaccgctc ccaacggccc    77340
ctcccagcag cgcgtcatcc gcgccgcgct ggacagcgcc cacctcaccg ccgccgacat    77400
cgacgccgtc gaggcccacg gcaccggcac cacccctcggc gacccgatcg aggcccaggc    77460
gctcctggcg acctacggac aggaccggcc gcggccctg tggctcggct cggtgaagtc    77520
caacatcggc cacacccagg ccgcctccgg tgccgccggc gtgatcaaaa tgatcatggc    77580
gttgcagcgc ggcgtgctgc cgcgcagcct gcacgccacc gaacccacca cggacgtcga    77640
ctggaccgcc ggctccgtcg acctcctcga cgagacggtc gcctggcccg agaccggacg    77700
cgcccgccgc gccggcgtct cctccttcgg catcagcggc accaacgccc acgtcatcct    77760
cgaacaggcc cccaccgccc ccgaagagcc caccaccgaa cccaccgtcc gccccgccgt    77820
cgtcccgtgg gcgctctccg cccgcaccgc cgccgccctc gacgcccagc gcgcccgcct    77880
caccggccac ctcgccgaca cccccgacgc cgacccctc gacgtcggct acgcgctcgc    77940
cgacggacgc gccaccttcg aacaccgcgc cgtcctgctc cccgacggca ccgaactcgc    78000
ccacggaacc gccggcgaag gccctgcgc cgtcctcttc tccggccagg gctcccagcg    78060
cccgggcatg ggacgcgaac tccacgcccg cttcccggtg ttcgccgcgg ccttcgacga    78120
gatcacagcg ctcctcgaca cccacctcga ccgcccgctg cgcgaggtcg tctggggcac    78180
cgacgccgac ctgctgaacg acaccggctg ggcccaaccc gccctgttcg ccgtcgaggt    78240
cgccctctac cgcctggtcg cgtccctcgg cgtcacccccc gacttcgtcg cggccactc    78300
catcggcgag ctcgccgccg cgcacgtcgc cggggtcctc tccctcgaag acgcctgcac    78360
cctcgtcgcc gcccgcgccc gcctcatgca ggccctgccg cgcggcggcg cgatgctcgc    78420
gatccgcgcc accgaggacg aggtcacccc ccacctcacc gacgacgtct cgatcgccgc    78480
cgtcaacggg cccacctccg tcgtcgtcgc cggcaccgag gaagccgtcg ccgcgatcgg    78540
ggcgcgcttc accgcccagg accgcaagac cacccggctg cgggtcagcc acgccttcca    78600
```

```
ctcgccgctc atggacccga tgctggcgga attccgcgcc gtcgccgcgg gcctgaccta   78660
ccacgagccg cgcatcccgg tcctctccaa cctcaccggc accgtcgccg ccgtcgccga   78720
cctgtgctcc gccgactact gggtccgcca cgtccgcgag gcggtccgct tcgccgacgg   78780
cgtcaccgcc ctcaccgacc gcggcgtgac cacgctcgtc gaactcggcc cggacggcgt   78840
gctgtccgcc atggcccagg aatccctgcc ggacggcgcc gccgccgtgc cgctgctgcg   78900
caaggaccgc cccgaggagc tctccgccgt caccggcctg gcccgcgccc acgtccgcgg   78960
cgtcacggtc cgctgggccg gcctcttcga cggcaccggc gcgcgccgcg ccgacctgcc   79020
cacctacccc ttccagcacc agcggttctg gccgaccgcg gcccgcgccg cccaggacgt   79080
caccgccgcg ggactgggcg ccgccgacca cccgctgctc ggcgccaccg tcgaactcgc   79140
cgacggggcc ggctacttgt tcaccagccg gctctccgtc cggacccacc cctggctcgc   79200
cgaccacggg gtccagggcc gggccctgct gcccggcacc gccttcgtcg aactggccgt   79260
ccgcgccggc gacgaggccg gctgcgaccg cgtcgaggaa ctgaccctgg ccgcccccct   79320
ggtgctgccc gagcgcggcg gcgtccaact ccaggtccgc gtcggcgccc ccgacgccgc   79380
cggccgccgc accctcggca tcttctcccg cgtcgaggac ggcttcgacc tgccctggtc   79440
gcaacacgcc accggcgtcc tgaccgccgg cgccggcgcc cccgacccca ccttcgacgc   79500
caccgtctgg cccccagcg gcgccgaacc cgtcgacctc accggcgcgt acgagcgcct   79560
ggccgcactc ggcttccagt acggccccgc cttccagggc ctgcgcgccg cctggcgccg   79620
cgacaccgag gtctacgccg aagtggccct gcccgacggc gcggacaccg accccgccgc   79680
cttcggactg cacccggccc tgctggacgc cgcacaacac gccgccgcct acgccgacct   79740
cggcgccatc agccgcggcg gcctgccgtt cgcctgggaa ggcgtctcgc tcgccgccgc   79800
cggcgccacc accgtccgcg cccggatcgc cccggccggc gaggacaccg tcaccatcgc   79860
cgtctacgac gccgccggcg gcaccgtgct gtccgtcgac tccctggtct cccgcgaggt   79920
ccccgccgac gcacccggcg ccgccggcac cgtccaccgc gactccctct ccacgtcga    79980
gtggaccccg ctccagggcc gcccgggccc cgcaccggcc accgtcgccg tcctcggccc   80040
cgacccggac gccctcgccg acaccctccg cgccaccggc atccggacca ccgccccccg   80100
cgacctggcc gccctcgccg acgccgaagg gcccgtcccc gacctggtcg tcaccaccct   80160
caccaccacc ccgggcgccc ccgtccccga cgccgcgcac gccaccaccg ccgccgtcct   80220
cgccctcgcc caacagtggc tcgccgacga ccgcttcgcc gacgcccgcc tggtcctcgt   80280
cacccgcggc gccaccgacg gcaccgaccc cgccgccgcg gccgccggcg gcctgatccg   80340
caccgcccgc accgagaacc ccggccgttt cgccctcctc gacctcgccc ccgacaccgg   80400
ccggcccgac cccgagaccc tggccaccgc cctggccgcc agccacgacg agcccgacct   80460
cgccgtccgc ggcaccgacg tgcacgccgc ccgcctggcc cgtgtcccgc tcgccaccga   80520
acccaccacc tggaacccgg acggcaccgt cctgatcacc ggcggcaccg gcggcctggg   80580
cgcggtcctc gcccgccacc tggtcgccac ccacggcgtc cgccacctgc tgctcgccag   80640
ccgccgcggc ccggccgccg acggcgccga cgacctgacg gccgaactca ccgggctcgg   80700
cgccaccgtc cacatcgccg cctgcgacgt cgccgacccg gccgccctcg ccgacctgct   80760
cggcaccgtc ccggccgggc acccgctcac cgtcgtcgtc cacaccgccg gcgtcgtcga   80820
cgacggcgtc ctcggctccc tcacccccgca gcgcctggac accgtcctgc ggcccaaggc   80880
cgacgccgcc tggcacctgc acgaggcgac ccgccacctc gacctggacg ccttcgtcct   80940
cttctcgtcc gtcgccgcca ccctcggcag ccccggacag gccaactacg ccgccggcaa   81000
cgccttcctg gacgccctcg ccgccggcg cgccgccacc ggcctgcccg ccacctccct   81060
cgcctggggc ccgtggaccc agagcgtcgg catgacaagc agcctgtccg acctcgacgt   81120
cgagcgcatc gcccgctccg gcatgccccc gctgaccctg gaacagggca ccgccctctt   81180
cgacgcggcc ctggccgccg ggcccgccgc cctcgccccg gtccgcctcg acctgcccgt   81240
cctgcgcacc cagggcgaca tcgccccgct gctgcgcggc ctgatccgca cccccgtgcg   81300
gcgcaccgcc gcccaggtct cgcagaccgc cgacggcctc gcccagcggc tcgccggcct   81360
cgacgccgcc gcccgccggg aagccctcct ggaactcgtc cgcacccaga tcgcccaggt   81420
cctcggccac gcggacgcca ccgaggtgga gaccggccgc cagttccagg acctcggctt   81480
cgactccctc accgccgtcg aactccgcaa cgccctgaac accgccaccg gcctgcggct   81540
gcccgccacc atggtgttcg actacccgac accacacgcc ctcgccgacc acctgcgcga   81600
cgaactcctg ggcaccgagg ccgagtcgac caccgccgtc cccgtgccga cccgtaccgc   81660
cggcaccgac gacccgatcg tcatcgtcgg catggcctgc cgctaccccg gcggcatcgc   81720
ctcacccgag gacctctggc gcctggtcag ccagggcgcc gacgccactg gcccgttccc   81780
caccaaccgc ggctgggacc tggacaacct ctacgacccc gaccccgacc gcccgggccg   81840
cacccacgtc gcgccggcg gcttcctgca cgacgccggc tccttcgacg ccgacttctt   81900
cgggatgagc ccgcgcgagg cgatggccac cgactcccag cagcgcctgc tgctcgaact   81960
ctcctgggaa gccgtcgaac gcgccggcat cgaccccgcc tcactgcgcg actccggcac   82020
cggcgtcttc gccggcgtca tgtacaacga ctacggcacc accctgaccg gcgacgagta   82080
cgaggcgttc cgcggcaacg gcagcgcccc gagcgtcgcc tccggccgcg tctcctacac   82140
cctcggcctg gaaggcccgg ccgtcacggt ggacaccgcc tgctcttcct ctctggtcgc   82200
```

```
cctgcactgg gcggcgcagg cgttgcgggc gggggagtgc tcgttggcgt tggccggtgg   82260
tgtgacggtg atgtcgacgc cgagcacgtt cgtggagttc tcgcggcagc ggggtctggc   82320
gcctgatggt cgttcgaagg cgttcgccga ggccgcggac ggcgtggcct ggtccgaggg   82380
cgtcggcatg ctggtcctgg agcggcagtc ggacgcggtg cgcaacggtc acgagatcct   82440
ggccgtggtg cgcggctcgg cggtcaacca ggacggtgcg tccaacggtc tgaccgcgcc   82500
caacggcccg tcccagcagc gggtgatccg tcaggcgttg gccagtggcg gcctgtccac   82560
ggccgacgtg gacgccgttg aggcgcacgg cacgggtacg acgctcggtg acccgatcga   82620
ggcccaggcg ctcctggcca cctacggtcg cgaccgcgac cccgagaacc cgctgctgct   82680
cggctcgatc aagtccaaca tcggtcacac ccaggcagcg gccggtgtcg ccggtgtcat   82740
caagatggtc atggcgatgc ggcacggcgt gctgccgcag accctgcatg tcgacgcgcc   82800
gtcctcgcac gtcgattgga gcgtcggcgc cgtcgaactg ctcaccgagc agaccgcctg   82860
gccggagacc ggccgggccc gtcgcgccgg tgtctcctcc ttcggcatca gcggcaccaa   82920
cgcccacgtc gtcatcgagc agtccccgac cgccgtcccc gccacgcccg cgtccgccga   82980
ccggtccgtc gaggaaccgc cggccgtccc ctgggccctg tccggcaaga cccccgacgc   83040
cctccgcgac caggccgccc gcctcctcgc ccacgtcgag gcccacccceg cactgcgccc   83100
cgtcgacatc agctactccc tgatcgccac ccgcaccgcc ttcgaccacc gcgccgtcgt   83160
cctcggcacc gaccgcgccg aggccctgcg cgccctcacc gccctcgccg ccggcgagac   83220
cgacccggcc gccctcaccg gcaccgtccg caccggccgc accgccttcc tcttctccgg   83280
ccagggctcc caacggctcg gcatgggccg cgtcctctac gagcggttcc ccgccttcgc   83340
cgaagccctc gacaccgtcc tcaccgccct cgacgcggaa ctcggccacc ccctccgcga   83400
catcatctgg ggcgaggacg ctcaactcgt cgaccggacc ggctacaccc aacccgccct   83460
gttcgccatc gaggtggcac tcttccgcct cctggaagcc tggggcatca caccggactt   83520
cgtggccggc cactccatcg gcgagatcgc cgccgcacac gtcgccggcg tgctctccct   83580
cggcgacgcc tgccgcctcg tcgtggcccg cgccgtgctg atgcagtcgc tgcccgaagg   83640
cggcgcgatg atcgccgtcc aggccaccga ggacgaggtc ctgcccctcc tcaccgacga   83700
cgtctcgatc gccgccgtca acagcccgac ctccgtcgtc gtctccggct acgagaacgc   83760
caccctcgcc gtcgcccggc acttcgccga ccagggccgc cgcaccacgc ggctgcgcgt   83820
cagccacgcc ttccactcgc cgctgatggc gccgatgctc gacgacttcc gcgccgtcgt   83880
cgagagcctc accttcaccg cccccacgac ccccgtcgtc tccaacctga ccggcgaact   83940
ggccccggcc gaggcgctct gctcggccga ctactgggtc cggcacgtcc gcgaggcggt   84000
ccgcttcgcc gacggcatcc gcaccctcgc cgaccgcggc gtcaccacct tcgtcgaact   84060
cggccccgac agcgtgctgt ccgccatggc ccaggagtcc gcccccgaag gcgccggcac   84120
catcccgctc ctgcgccgcg accggcccga ggaacaggcc gtcctggccg ccctctgcca   84180
cctccaggtg ctcggcgtcg aggccgactg gtccgccacc ttccgcggcc tcgaccccgt   84240
ccgcgtcgac ctgccgacct acgccttcca gcaccgctgg ttctggcccg ccgcccgacc   84300
cgcccgcccc gacgacgtcc gcgccgccgg cctgggcgcc gccgaacacc ccctcctcgg   84360
cgccgccgtg caactccccg acgacgacgg cgcactcttc accggccgcc tctccctgcg   84420
cacccacccg tggctggccg accacaccgt cctgggcacc gtcctgctcc cgggcaccgc   84480
actggtggaa ctcgccgtcc gcgcgggcga cgagaccggc agcggccacc tcgaagaact   84540
caccctcgcc gcgcccctga ccctccccga ggacggcgcc accctcctcc aggtccgcgt   84600
cggatccgcc gacgacaccg gccgccgcac cgtcaccgtc cacgcccgcc ccgacgacac   84660
cgccgaccgc acctggacgc tgcacgccac cggtgtgctc gccaccacgc caccggccgc   84720
cgcggcgttc gacaccacgg tctggccgcc cgccgacgcc gaacccctca ccaccgacga   84780
ctgctacgca cacttcacca cccaccgctt cgcctacggc cccgccttcc agggcctgcg   84840
ggccgcctgg cgcgccggcg acgtgctgta cgccgaggtc gccctgccgg agtccgccac   84900
cgacgaagcg gccgccttcg gcctgcaccc ggcgctcctg gacgccggcc tgcacgccgc   84960
gctcctcgcc gacgaccgcg acaccggact cccgttctcc tgggaaggcg tcactctgca   85020
cgcctccggc gccaccgcgc tacgcgtccg gctcgccccg aacggcccca acggcctgtc   85080
cgtcaccgcc gccgacccgg ccggcaaccc cgtcgccacc gtcacccgcc tgctcgcccg   85140
cccccctggac gccgagcagt tgaccatcca cagcgccctg acccgcgacg cgctcttcca   85200
cctggactgg accccggtcc cgcttcccga caccgccaac tccgcgccgc cggccctcct   85260
cggcccggac accgccgtgc tcgccgacgc cctcggcgac ccggccgtcg cacgccacgc   85320
aaccctcgac gacctcctgg ccggggacac cacccegccc gccacggtcc tcgtccccct   85380
cggcgcccca ctcgacggcg acaccgcgca gcacgcgcac gccctcaccc gcagcgcgct   85440
gaccctcgtc cagcagtggc tcgccaccga ccgcctcgcc gactcccgcc tggtcttcgt   85500
cacccacgga gccgtcgcca ccgacgacgc gcccccacc gacctggccg ccgccgcggt   85560
ctggggcctg atccgctccg cgcagaccga gaaccccggc accttcaccc tcctcgacct   85620
cgacaccgag cccgactcga ccaccgcgct cagccgcgcc ctgaccctcg acgaaccaca   85680
gctcctcctc cgcgccggcc gcgccgcgc cgcccgcctc acccgcaccc cgcccccac   85740
caccaccacc cacacgccgt ggtccgcgga cggaacggtg ttggtgacgg gtggtacggg   85800
```

```
tggtctgggt gggttggtgg cccggcatct ggtgcggtcg tgtggggtgc ggcatttgtt    85860
gttgaccagt cgttctggtg tgggtgctgc gggtgcggcc gggttggtcg cggagttgga    85920
gtcgttgggc gcgcgggttg tggttgcggc gtgtgatgtg ggtgatggct cggctgttgc    85980
ggagttggtt gccggtgtgt cggagtcgta tccgttgtct gcggtggtgc atgcggctgg    86040
tgtgttggat gacggtgtgg tgggttcgtt gacgccggag cggttggctg cggtgttgcg    86100
tccgaaggtg gatggtgcgt ggaacctgca tgaggcgacg cgtggtctgg atctggacgc    86160
gtttgttgtc ttctcgtctg ttgcgggtgt gttcgggggt gcgggtcagg ccaactatgc    86220
ggcgggtaat gcgtttttgg acgcgttgat ggttcatcgg gtggctggtg ggttgcctgg    86280
tgtgtcgttg gcgtggggtg cttgggatca gggtgtgggg atgacggcgg ggctgacgga    86340
gcgggatgtc cgtcgtgctg ctgagtcggg tatgccgttg ttgacggttg atcagggtgt    86400
ggcgttgttc gatgcggcgt tggcgacggg gagtgccgcg ttggtgccgg tccgtctgga    86460
cctggccgca ctgcgcaccc ggggcgacat cgcaccgctc ctccgcggcc tcgtccgcgc    86520
accgctgcgc cgcaccgcgg ccaccggcct cgccaccggc gcggacaccg gcctcgtcca    86580
acggctcggc cgactcgacc acgcccaacg ccacgaggca ctgctcgaca tggtccgcag    86640
cagcgccgcg ctcgtcctcg gccacgccga cggcaacgcc atcgacgccg aacgcgcctt    86700
ccgcgacctc ggcttcgact cgctcaccgc ggtcgaactc cgcaaccgtc tgcgcaccgc    86760
caccggcctg cacctgtcgg ccaccatggt cttcgaccac cccaccctgt ccgccctcgc    86820
ggagcacctg cgggacgagt tgttcggcgc ggtcgagagc gaggtgcggg tgccggtcca    86880
ggcactgccg ccgaccgccg acgatcccat cgtggtggtg ggcatggcct gccgtttccc    86940
cggtggtgtg acctcgcccg aggacctgtg gcgcctggtc gacgacggca ccgacgccat    87000
caccaccttc ccgaccaacc gcggctggga cctggacaac ctctacgacc cggaccccga    87060
gcacttcggc acgtcgtaca cccgctccgg tggcttcctg cacgaggcgg gggagttcga    87120
cccggcgttc ttcggaatga gcccgcgtga ggcgctggca accgactccc aacagcgtct    87180
cctgctggaa tcctcctggg aggcgatcga gcgggccggc atcgacccgc tgaccctgcg    87240
cggcagcgcc accggcgtct cgccggcgt gatgtacagc gactacggga gcatcctcgg    87300
cggcaaggag ttcgagggct tccaaggcca gggaagtgcg ggcagcgtgg cctcgggccg    87360
cgtctcctac gccctcggct tcgagggccc ggccgtcacg gtggacacgg cttgctcttc    87420
ctctctggtc gccctgcact gggcggcgca ggcgttgcgg gcgggggagt gctcgttggc    87480
gttggccggt ggtgtgacgg tgatgtcgac gccgagcacg ttcgtggagt tctcgcggca    87540
gcggggtctg gcgcctgatg gtcgttccaa ggcgttcgcc gaggccgcgg acggcgtcgg    87600
ctggtccgag ggcgtcggca tcctcgtcct ggagcgccag tcggacgcgg tgcgcaacgg    87660
ccacgagatc ctcgccgtga tccgcggctc ggcggtcaac caggacggtg cgtccaacgg    87720
cctgaccgcg cccaacggcc cgtcccagca gcgcgtcatc cgtcaggcgt tggccagtgg    87780
cggcctgtcc acggccgacg tggacgccgt cgaggcgcac ggcacgggta cgacgctcgg    87840
tgacccgatc gaggcccagg cgctcctggc cacctacggc cgtgaccgcg accccgagaa    87900
ccccctgtgg ctgggctccc tgaagtccaa catcgggcac acccaggcag cggccggtgt    87960
cgccggtgtc atcaagatgg tcatggcgat gcggcacggc gtgctgccgc agaccctgca    88020
tgtcgacgcg ccgtcctcgc acgtcgattg gagcgtcggc gccgtcgaac tgctcaccga    88080
gcagaccgcc tggccggaga ccggccgggt ccgtcgcgcc ggtgtctcct ccttcggcat    88140
cagcggcacc aacgcccacg tcatcgtgga acagccggcg ctcgtcgaaa gcccggccgc    88200
ggagccgagc ggacgcgaac ccggcgtcgt tccgctgccg ctgtccggaa agtcccccga    88260
ggccctgcgc gaccaggccg cacgcctgct ggccgggttg gcggagcggc ccgcgctgcg    88320
cccgctcgac ctcggctact cgctggcgac gacccgttcg gcgttcgacc accgggcggt    88380
ggtgctcgcc accgaccgcg ccgatgcggt ccgcgcgctg acggcgctcg ccgccgccga    88440
cgcggatctc tccgccgtcg tcggcgacac ccgcacgggt cgtcacgcgg tgctgttctc    88500
gggtcagggc tcgcaacgcc tgggcatggg gcgtgagttg tacgagcgtt ccccggtctt    88560
cgccgaggct ctcgatgtcg cgatcgacca cctggacgcc gccttgcccg cccaggccag    88620
tctgcgtgag gtgatgtggg gcgacgatgt cgagctgctg gacgagacgg gttggacgca    88680
gccggctctg ttcgccgtcg aggtggccct gttccggctg gtggagagtt ggggtgtccg    88740
tccggacttc gtggccggtc attccatcgg tgagatcgcg gcggcgcatg tcgtcggggt    88800
gttctcgctg gaggacgcct gccgtctggt ggccgcccgt gcgacgctga tgcaggcgct    88860
gccgaccggc ggcgcgatga tcgcgatcca ggccgccgag gacgaagtca cccagcacct    88920
gactgacgac gtctcgatcg ccgccgtcaa cggcccgacc tccgtggtcg tctccggggc    88980
cgagagcgct gcccgcacgg tggccgaccg gctcgcggag aacggccgca agacgacccg    89040
gctgcgggtc tcgcatgcgt tccactcgcc gttgatggat ccgatgctgg cggagttccg    89100
tgcggtggcc gagggcctgt cctacgccac cccgaccctc cccgtcgtct cgaacctgac    89160
gggccggctg ccacggccg atgacctctg ctcggccgag tactgggcgc gccacgtccg    89220
cgaggcggtc cgcttcgccg acggcgtcag caccctggag aacgagggcg tcaccacgtt    89280
cctggaactg ggaccggacg gcgtgctgtc cgccatggcc cagcagtcgc tcaccggcga    89340
cgccgccacc gtcccggccc tccgcaagga ccgcgacgag gagacgtccg cgctcaccgc    89400
```

```
cctcgcccac ctccacacgg caggtctccg cgtcgactgg gcggcgttct tcgccggcag   89460
cggcgccacc cgcgtggacc tgccgaccta cgccttccag cacgccacct actggcccac   89520
cggcaccctg cccaccgcgc acgccgcggc cgtcggcctc accgccgccg agcacccgct   89580
gctgaacggt tcggtcgaac tcgccgaagg cgaaggggtg ttgttcaccg gacggctgtc   89640
actgcagtca catccgtggc tggccgacca cgccgtcatg ggacaggtcc tgctgcccgg   89700
caccgcactg ctggaactgg cgttccgggc cggcgacgag gccggttgcg accgcgtcga   89760
ggaactgacg ctcgccgcac cgctcgtcct gcccgagcgc ggtgcggtac agacccaggt   89820
ccgggtcggc gtcgccgacg acaccggccg ccgtaccgtc accgtccact cccggcccga   89880
gcacgcgacc gacgtgtcgt ggacccagca cgcgaccggc accctgacca tgggctccgc   89940
cccggccgac accggtttcg acgccactgc ctggccgccc gccgacgccg aacccctcgc   90000
caccgacgac tgctacgcgc gcttcacgac gctcggcttc gcctacgggc cggtcttcca   90060
gggcctgcgg gccgcctggc gcgccggtga cgtgctgtac gccgaggtgg ccctggcgga   90120
gtccaccggc gacgaggcga ccgccttcgg tctgcacccc gcactgctcg acgccgccct   90180
gcacgcctcc ctcgtcgccc acgagggcga ggagagcaac ggcggactgc cgttctcctg   90240
ggagggcgcg accctctacg cgaccggcgc caccgcgctg cgcgtccggc tgaccccgac   90300
gggcaccgac ggccgttcgg tggccatcgc cgtggccgac accgccggtc gtccggtcgc   90360
cgccatcgac aacctcgtct cgcgccgggt ctccggcgac cagttgaccg gcgccgcggg   90420
actgcccgc gacgccctgt tcaccctgga ctggaacccc gtaccggaga acctcgtacc   90480
ggagaaccct gtaccggaga acaccggcgg gggccacgcc caggaccagg acggccggcc   90540
cgccgcggcc accgtcgcgc tggtcggcgc ggacggcacc gcgatcgccg ccgacctgac   90600
cgccgccggc atccacacca ccctccaccc cgacctcacc accctcgcca cgaccgacgc   90660
cgacgttccg aagacggtcc tcatccccct caccggaacc ggaaccggaa ccggcaccgg   90720
gactgagtcg acggacggaa tcgggacggg ggccgccgag tcggacgcgt ccgcccccctc   90780
cccggccgag gtcgcccaca ccctgtccac cgccgcactc gccctcgtcc aggagtggac   90840
cgcacaggag cgcttcgccg gctcccgcct ggcgttcgtc acgaccgggg cgacggccgc   90900
cggcggtacc gacgtcatgg acgtggccgc cgccgcggtc tggggcctgg tccgatccgc   90960
ccagtccgaa gccccggaca ccttcgtcct gatcgaccgt gacccccggcc cggccggcac   91020
gcacgaccgc acagccgccg ccgaacgggg ccaactgctc ctacgggcac tgcacaccga   91080
cgaaccgcag ctcgccctgc gtgacggcgg cgtgctcgcc gcccgcctgg cccgcttcga   91140
caccgcggcc gcgctcaccc cgccggccga ccgggcctgg cggctcgaca gcacggccaa   91200
gggcagcctc aacggcctcg ccctgacccc gtatccggcg gcactggcgc cgctcaccgg   91260
ccacgaggtg cgggtcgagg tgcgtgccgc gggcctgaac ttccgtgacg tgctcaacgc   91320
gttggggatg tatccgggtg atgatgtcgg atcgttcggt tcggaggcgg ccggtgtggt   91380
cgtcgaggtc ggaccggagg tgaccggcct ggccccggc gaccaggtca tgggcatgat   91440
caccggcagc ttcggctcgc tcgccgtgga cgacgcgcgg cgcctcgccc gcctgcccga   91500
ggactggtcc tgggagacgg gtgcgtcggt gccgttggtg ttcctcaccg cgtactacgc   91560
cctgaaggag ttgggtggtc tgcgggcggg ggagaaggtg ctggtgcatg ccggtgccgg   91620
tggtgtcggt atggcggcga tccagatcgc ccggcatgtc ggtgccgagg tgttcgccac   91680
ggccagtgag ggcaagtggg acgtgttgcg ctcgctcggc gtggccgacg accacatcgc   91740
ctcctcccgc accctcgact tcgaggcggc cttcgccgaa gtcgccggcg accgcggcct   91800
ggacgtcgtc ctcaactccc tcgccggtga cttcgtcgac gcctcgatgc ggttgctcgg   91860
cgacggcggc cggttcctgg agatgggcaa gaccgacatc cgcgccgcgg actccgttcc   91920
cgacggcctc tcctaccagt ccttcgacct cgcctgggtg gtgccggaaa ccatcggcac   91980
catgctggcc gagctgatgg acctcttccg caccggcgca ctgcggccac tcccggtccg   92040
cacctgggac gtccggcacg ccaaggacgc gttccgcttc atgagcatgg ccaagcacat   92100
cggcaagatc gtgctcaccc tgccccgctc ctggaagccc gagggaacgg tgttggtgac   92160
gggtggtacg ggtggtctgg gtgggttggt ggcccggcat ctggtgcggt cgtgtggggt   92220
gcggcatttg ttgttgacca gtcgttctgg tgtgggtgct gcgggtgcgg ccgggttggt   92280
cgcggagttg gagtcgttgg gcgcgcgggt tgtggttgcg gcgtgtgatg tgggtgatgg   92340
ctcggctgtt gcggagttgg ttgccggtgt gtcggagtcg tatccgttgt ctgcggtggt   92400
gcatgcggct ggtgtgttgg atgacggtgt ggtgggttcg ttgacgccgg agcggttggc   92460
tgcggtgttg cgtccgaagg tggatggtgc gtggaacctg catgaggcga cgcgtggtct   92520
ggatctggac gcgtttgttg tcttctcgtc tgttgcgggt gtgttcgggg gtgcgggtca   92580
ggccaactat gcggcgggta atgcgttttt ggacgcgttg atggttcatc gggtggctgg   92640
tgggttgcct ggtgtgtcgt tggcgtgggg tgcttgggat cagggtgtgg ggatgacggc   92700
ggggctgacg gagcgggatg tccgtcgtgc tgctgagtcg ggtatgccgt tgttgacggt   92760
tgatcagggt gtggcgttgt tcgatgcggc gttggcgacg gggagtgccg cgttggtgcc   92820
ggtccgtctg gacctggccg cactgcgcac ccggggcgac atcgcaccgc tcctccgcgg   92880
cctggtcaag gcgcccatcc gccgcgcagc cgccaccaca cccggcgaca ccggactcgc   92940
cgagcagctc acccggctcc agcgcgccga gcgacgggac accctcctcg cgctcgtccg   93000
```

```
cgaccaggcc gcgatggtcc tcggccacac ctcgggcgac ggcgtcgacc cgtcccgcgc   93060
cttccgcgac ctcggcttcg actcgctcac cgcggtcgaa ctccgcaacc gcatcggcgc   93120
ggccaccggc ctgcggctac cggccacggc cgtcttcgac taccccaccg ccgatgccct   93180
cgccgcacac ctgctcaccg aactgctcgg ccccgacgcc gagtcggacc ccgacgagcc   93240
cggcgacccc accgcgggac cgaccgacga ccccatcgtc atcatcggca tgagctgccg   93300
cttccccggc gacatcggct cgccggagga cctgtggcgc ctgctcggcg acggcgccga   93360
cgtcgtcacc gacttcccga ccaaccgcgg ctgggacctg gacaacctct acgaccccga   93420
ccccgcgcac gccggcacct cgtacgcccg caccggcggt ttcctgcacg acgccgccga   93480
cttcgacgcc gacttcttcg gcatgagccc ccgcgaggcc atggccacgg actcccagca   93540
gcgcctgctg ttggagtcct cgtgggaggc gatcgagcgg gccggcatcg acccgctgac   93600
cctgcgcgac agccgcaccg gcgtcttcgc cggcgtcatg tacagcggct acggcacccg   93660
cctcgacggc gccgaattcg aaggcttcca ggggcagggc agcgcactga gcgtggcctc   93720
cggccgggtc tcctacacct tcggcttcga aggcccggcc atgacggtcg acaccgcctg   93780
ctcctcctcg ctggtcgccc tgcacctcgc cgcacaggca ctccgcggcg gtgagtgcac   93840
cctcgccctc gccggtggtg tcaccgtgat gtccatcccg gacaccttca tcgagttctc   93900
ccggcagcgc ggactggccc ccgacggccg ctccaagccg ttctccgagt ccgccgacgg   93960
cgtcggctgg tccgagggcg tcggaatgct gctcctggag cgccagtcgg acgccgtgcg   94020
caacggccac cagatcctgg ccgtggtgcg cggctcggcg gtcaaccagg acggtgcgtc   94080
caacggcctg accgcgccca acggcccgtc ccagcagcgg gtgatccgtc aggcgttggc   94140
cagcggcggc ctgtccacgg ccgacgtgga cgccgtcgag gcgcacggca cgggcaccac   94200
gctcggtgac ccgatcgagg cccaggccct cctggccacc tacggccgcg accgcgaccc   94260
cgagaacccg ctgctgctcg gttcgatcaa gtccaacctc ggccacaccc aggcagccgc   94320
cggtgtcgcc ggcgtcatca agatggtcat ggcgatgcgg cacggcgtgc tgccccgcag   94380
cctgaacatc accgagccgt cctcgcacgt cgattggagc gccggcgccg tcgaactgct   94440
caccgagcag accgcctggc cggagaccgg ccgggcccgt cgcgccggta tctcctcctt   94500
cggcatcagc ggcaccaacg cccacgtcat cctggagcag ccggaggccg cgcggcactc   94560
ggcgccggaa gaagccgaca cggcggaggc agccgccaag gcgccggcca ccgcgcacct   94620
gcccgtaatg ccgtgggcac tgtccggcaa gacgccggag gccctgcgtg cccaggccgc   94680
acgcctcctc gcccacctcc agcagcgccc cgaactcgca cccgccgaca tcgccctgtc   94740
cctcgccacc cagcgctccc agttcacccca ccgggcagtc gtcctgagca ccgaccgtga   94800
cgaggcgacc cgcgcgctgt ccgccctcgc caccaccgcc gcgtccgacc cctcggccct   94860
caccggcacg gtcaccatgg gacgttgcgc ggtgctgttc tcgggtcagg gctcgcaacg   94920
tctgggcatg gggcgtgagt tgtacgagcg tttcccggtc ttcgccgagg ctctcgatgt   94980
cgtgatcgat cacctggacg ccgccttgcc cgcccaggcc ggtttgcgtg aggtgatgtg   95040
gggcgacgat gtcgagttgc tgaacgagac gggttggacc cagcccgcgc tcttcgccat   95100
cgaggtggcg ctgtttcggc tggtggagag ttggggtgtc cgtccggact tcgtggccgg   95160
tcattccatc ggtgagatcg cggcggcgca tgtcgtcggg gtgttctcgt tggaggacgc   95220
gtgccgtctg gtggccgcgc gggcgacgct gatgcaggcg ttgccggccg gtggcgcgat   95280
gatcgcggtc caggcgaccg aggacgaagt catcccgcac ctgaccgacg aggtggcgat   95340
cgcggccgtc aacggcccga cctccgtggt gatctcgggc gcagaagagg ccacgcagac   95400
cgtggcacaa cacttcgccg accaggggcg ccggacgacc gcgctgcggg tctcgcatgc   95460
gttccactcg ccgctgatga tgctggcgga gttccgtgcg gtggccgagg gcctgtccta   95520
cgccaccccg accctccccg tcgtctcgaa tctgacgggc caggtggcca cggccgacga   95580
actctgctcg gccgagtact gggtgcgcca cgtccgtgag gcggtccgct tcgccgacgg   95640
tgtgacggcc ctcgaagccg agggcgtgcg gaccttcctg gaactcggcc cggacggcgt   95700
cctcgccgcc atggccaggg aaaccgtcgc cgacgacacg gtcaccgtcc ccgtcctccg   95760
caggaacatg cccgaggaac ggaccctgct caccgcactc ggccggctcc acaccaccgg   95820
aaccccgatc gactgggccg ccctcctggc cccgaccggc gcccgccgg tggacctgcc   95880
gacatacgcg ttccaacacc gtcccttctg gccctccggc ccccgcgaca ccgcggatgc   95940
cgccgccgtc ggcatcgccg gcgcgagcca cccgctcctc aacggcatcg tcgaactcgc   96000
cgacgaagag ggcctgttgt tcaccggacg gctgtcactc cagtcgcatc cgtggctggc   96060
cgaccacgcc gtcatgggac aggtcctgct gcccggcacc gcactgctgg aactcgccct   96120
gcgcgccggc gacgaggtcg gctgtgacca cgtcgaggaa ctgacgctcg ccgcaccgct   96180
cgtcctgccc gagcgcggcg cggtacagac ccaggtccgg gtcggcgtcg ccgacaccac   96240
cgggcgccgc accgtcacga tccactcgcg tcccgcacgc gccacgacca ccgacagtga   96300
cacccacacc ggcaccgaca ccccgtggac ccaacacgcc accggcgtcc tcgtcgccgg   96360
cctgccggcg acggcaaccg tcccgttcga tgccaccgtg tggccgccgg cgcacgccga   96420
acccgttgac ctggcggact tctacgcgtc ccgggccggc gaaggattcg gctacgggcc   96480
cgctttccag ggcctgcgag cggcctggcg ccgcgacggc gaggtgttcg ccgatgtcgc   96540
actgccggag gccggccgta ccgaagccga ggcgtacggg ctgcatccgg cactgctcga   96600
```

```
cgccggactg cacgcagcct ggctcgtcgc cccggacggg gagcccacac ggacgggcag    96660
cgtgccgttc tcgtggcgcg gcgtcttcct ggccgcttcc ggtgcctcct cggtccgcgt    96720
ccgactcggc cgcgactccg acggaacgct gagcctggcc atcgccgaca ccaccggtgc    96780
accggtcgcg tccgtacagg ccctctccat gcgcaccgtc tcggtgacgg ccctgagcgc    96840
caccgcgggc ctcgcccgcg acgcgctgtt ccgcctggac tgggcctcgg ccccggagcc    96900
ggcgtgccag ccggacgaca cggtgaccgt gatcccggcg gtcgcggtcg tcggtacgga    96960
aacctccgaa ctcacctccg agctcaccgc ggccctgcgt gcggcgggcg ccgacgtcga    97020
cgtccgcacg accctgtcga ccgatgaacc cgcgcccgcg ctgatcgcgc tgccgctcgt    97080
cgcctccgac cagaccggca ccgcagaagc ggcacccgtc ccggcggccg tgcacgacct    97140
cacccgccga gccctggccc tcgtacagac ccgcctgcaa gagcagcact cgcggacac     97200
gaagttcgtc ttcgtgaccc gtggtgcgac ggtcgggcgt gatgtggctg ctgctgcggt    97260
gtggggtctg gtgcgttcgg cgcagtcgga gaatccgggt tgttttgctc tggtcgatct    97320
ggatccggat ggtgcggtgg gtgcggctgc gctcgtcgct gcgttggtca gtggtgagcc    97380
gcagcttgcg gtgcgcggtg atgtgttgcg ggtcgcgcgt ctggtgcggc ggccgctcac    97440
cgaggtcggt gcgggtgctg atggcaccgg ggatggcgtc gggggtggct ctggtgtgtc    97500
gttctcgggt gagggtgcgg tcctggtcac tggtggtacg ggtggtctgg gtgcggtgtt    97560
ggcgcgtcat ctggtggccg agtatggggt gcgggatctg ctgttggtca gtcgcagtgg    97620
tgaacgtgcc gtgggtgctg gggagttggt ggcggagctt gcgggtgtgg gtgcgcgggt    97680
gcgggtggtt gcgtgtgatg tgaccgatcg tgccgcggtg gtggagttgg ttggcgggca    97740
tgcggtgtcc gcggtggttc atgcggctgg tgtgctggat gacggcatgg tgggtgcgtt    97800
gaccggggag cggttgtccg cggtgctgcg gccgaaggtg gatgctgtct ggcatctgca    97860
tgaggcgacc cgcggcctgg atctggacgc gttcgtcgtc ttctcctctc tcgccggggt    97920
cttcggcagt cccggccagg ccaactacgc ggccgcgaac gccttcctgg acgcgctgat    97980
gacgcggcgc cgggcggagg gactgcccgg cctgtcactc gcatggggac cgtggtcgct    98040
gaccgatggc acatcgggca tgctcgcgga cgccgaggcc gatcgcctga cccgttcggg    98100
agtgccaccg ctgaccgcgg agcaaggact ggcactgttc gacgcggccc tggcgaccgg    98160
tgacgccacc tgcgtcccgg tccgcctgga cctgtcggca ctgcgtgccc agggtgaggt    98220
gccgcccttg ctgcggtccc tgatccgagg ccgctcgcgc cgcgccgccg ccgcggaatc    98280
ggcaaccgcc accgggctgc gggaacggct cgtcggactg aacccggtcg agcgacagga    98340
agtcctcctg gacctcgtgc gcggccaggt ggccctggtc ctcggccacg ccgacgccga    98400
cgacgttcat ccggcacgtg ccttcaggga gttgggcttc gactccctca cctcggtgga    98460
actgcgcaac cgcctcaaca ccgtcaccgg tctgcggctc cccgcgacca tggtgttcga    98520
ctatccgacc gtcgaagtgc tggtctccta cgtcctggac gagttgctgg cacggatgc     98580
cgaggtggcg accgtgcagc cggccgcggt tgcggtggcg gacgatccga tcgtcatcgt    98640
gggcatggcc tgccgctacc ccggcggggt cgcctctccg gacgacctgt ggaggctcgt    98700
cacggacggc gtcgacgcgg tgtccccctt cccgaccaac cgtggttggg acgtcgaatc    98760
cctctatcac ccggaccccg accatctcgg tacctcctac acgcgctcgg gtgggttcct    98820
gcatgaggcg ggggagttcg atccgggggtt cttcgggatg agtccgcggg aggcgttggc    98880
gaccgattcc cagcagcggt tgttgttgga gtcgtcgtgg gaggcgatcg agcgggccgg    98940
tattgatccg gtgagtttgc ggggtagtcg gacgggtgtg ttcgcggggg tgatgtacag    99000
cgattacagc gcgatgttgg cgagtccgga gttcgagggt ttccagggca gtgggagttc    99060
gccgagtttg gcgtcgggtc gggtggccta cacgttgggg ttggaaggcc cggcggtgac    99120
ggtggatacg gcgtgttcgt cgtcgttggt ggcgatgcac tgggcgatgc aggcgttgcg    99180
tagtggtgag tgtgggttgg cgttggccgg tggtgtgacg gtgatgtcga cgcctgcggt    99240
gtttgtggac tttgctcggc agcggggttt gtcgccggat ggtcggtgca aggcgtttgc    99300
ggatgcggcc gatggtgtgg gctggtccga gggcgtcggc gtgttggtcc tggagcggca    99360
gtcggacgcg gtgcgcaatg gtcacgagat tttggctgtg gtgcggggtt cggcggtcaa    99420
ccaggatggt gcgtccaatg gtttgacggc gcccaatggt ccgtcgcagc agcgggtgat    99480
ccggcaggcg ttggccagtg gtggtctgac ggccggtgac gtggatgtgg tggaggcgca    99540
tggtacgggt acgacgctcg gtgatccgat cgaggcgcag gcgttgttgg cgacgtatgg    99600
gcgggatcgt gagcctgagc ggccgttgtt gttgggttcg gtgaagtcga atctggggca    99660
tacgcaggct gctgcgggtg tggcgggcgt tatcaagatg gtgttggcga tgcggcatgg    99720
tgtggtgccg cggacgttgc atgtggatgc gccttcttcg catgtggact ggtccgaggg    99780
tgcggtggag ctgctcagtg agcaggcggc ttggccggag acgggtcggg tgcggcgggc    99840
gggtgtctcc tccttcggca tcagcggcac caacgcgcac gtgattctgg agcgtccgga    99900
ggccgcgcgg cgtccggtga tggagacgaa caccgtggag ccgtccacgg tgccgtgggt    99960
gctttccggc aagacaccgg aggccctgcg cgcccaggcg gcgaaactgt tgtcgtcgat   100020
cgaggaacgc ccggagcttc gcctcgtcga cgtgggaatg tccctggtca ccggccgctc   100080
gacctttgaa caccgtgcgg tggtactcgc cgccgaccgt ccgacgccg cccgcgcatt    100140
gtcggcgatt gctgccgacg aggccgatgc tgccgctgcc accggccgcg tcggcgcggg   100200
```

```
tcgtcacgcg gtgctgttct cgggtcaggg tgctcaacgt ctgggcatgg ggcgtgagtt    100260
gtacgagcgt ttcccggtct tcgccgaggc tctcgatgtc gtggtcgacc acctggacgc    100320
cgccttgccc gcccaggccg gtttgcgtga ggtgatgtgg ggcgacgatg cggagttgct    100380
gaacgagacg ggttggacgc agccggctct gttcgccatc gaggtggcgc tgttccggct    100440
ggtggagagt tggggtgtcc gtccggactt cgtggccggt cattccatcg gtgagatcgc    100500
ggcggcgcat gtcgccgggg tgttctcgct ggaggacgcc tgccgcctgg tggccgcccg    100560
tgcgacgctg atgcaggcgt tgccggccgg cggtgcgatg atcgcggtcc aggcgaccga    100620
ggacgaagtc accccgcatc tgaccgacga cgtagcgatc gccgccatca acgggccgaa    100680
cgcactggtc gtatcgggtg tggaggatgc cgccgtcgag atcggggcgc ggttcgcggc    100740
cgaggggcgt cgcacgaccc gactccatgt gtcgcatgcg ttccactcgc cgttgatgga    100800
tccgatgctg gcggagttcc gtgtggtggc ggagggcctg tcctatgctg ctccgtccct    100860
ccccgtcgtc tcgaatctga cgggccaggt ggccacggcc gacgaactgt gctcggccga    100920
gtactgggtg cgccacgtcc gcgaggcggt ccgcttcgcc gacggggtga cggccctcga    100980
agccgagggc gtgcggacct tcctggaact cggcccggac ggcgtcctcg ccgccatggc    101040
aggagcctcg ctcaccgaat cctccctcgc ggtaccgctg ctccgtaagg accggccgga    101100
ggaaccggcg gcactcgccg ccctggccca gttgcacatc gccggcgcgc gcgtcgattg    101160
gcccgtgctc ttcgctggtg tgggtgcggg gcgggtggag ttgccgacgt atgcgttcca    101220
gcgtgggtgg ttctggccgg ttggtcgggt tggtgttggt ggtgatgtgg gtgctgtggg    101280
gcttgggtct gcggggcatc cgttgttggg tgctgcggtg gagttggctg cgggtgcggg    101340
ggtggtgttg acgggtcgtc tgtcgttgtc gtcgcatggt tggttggctg atcatgcggt    101400
gatggggcgg gtgtttgttc ctggtacggc gttgctggag atggtgatgc gtgctggtga    101460
tgaggtgggg tgtggtcgtg ttgaggagct gacgttggcg gcgccgttgg tgttgcctga    101520
gcgtggtggg gtgcgggttc aggttgctgt ggatgctcct gatgctgcgg gtcgtcgtgg    101580
tgtggggggtg tattcgtgtc ctgatggtgt gggtcaggcg gtgtggtcgc agcatgctgt    101640
cggtgtgttg gcctctggtg tggctgacca ggtcggtggg ttcggtgacg gtggtgtgtg    101700
gccgccgcag ggtgcggtgt cggtggatgc tgagggctgc tacgagctgt ttgcggatgc    101760
tgggttcggt tatggcccgg tgttccaggg gttgcgtgcg gtgtggcgtc gcggcgagga    101820
actcttcgca gaggtcgccc tgtcggacga ggttgctgag agcgctgata cggcgaccgg    101880
tttcgggttg caccggccgt tgctggatgc ctcgctccat gcctcgcttc tctcctccct    101940
tgaaggtcaa tccgccgatg gtgggcctgc gttgccgttc gcgtgggagg gtgtttccct    102000
cttcgcctcg ggtgcgacgg ctttgcgcgt gcggttggcg ccggcgggtg agcatgcggt    102060
gtcggtgacc gcggtggatc cgaccggtgc gccggtgatt tccatcgacg cgcttcgtac    102120
ccgtcgcctc accctcgatg aggtcaacgc atctcacacc cagctgagcg atgcgctctt    102180
cggcgtccaa tggaccacgg tcccgagcac cccggccgcc gaccacccgt cggtagccat    102240
catcggaacc gaccacctgg ggctcgccga agcgctcagc agttcctctg ctggcgcgac    102300
cacgaccacg accgcggccg cgtacgagag ccttgacgcg ctgatcgcgg cggggcccga    102360
agtgtccgtc cctgacgtca cactcatcgg tctcaccacc gaggacgcca tcgctcagta    102420
cgtgaacgac cacgacgcca cggtggccgg ccaaggcacc atcggagccg cgcgcggcggc    102480
cgtagacgcg gctcgccgcc tcaccgccga ggccctgcgc acgatccagg catggttggc    102540
cgacgagcgt ctggcggcaa ggcgcctggt cttcgtgacc cgcggcgcgg ccgacgggca    102600
ggatgtcgcc gcggcggccg tccagggcct ggtgcgctcc gcccagaccg agaacccggg    102660
caccttcggc ctcctcgacc tcgacgggac cgaggcgtcg accgcggtcc tcggcgaggc    102720
tctcacctcc gacgaaccgc aactgcttct gcgcgacgga cacctgcacg ccgccaggct    102780
gacccgcctg gcgtcgcccg ccgacacggc cgtgcccacg gagtggaacg cggacggcac    102840
ggtgctgatc accggtggta ccggcgggct cggcgcgcag ttcgcacggc acctcgtcga    102900
caggtacggc gtccgcaatc tcctgctcgt cagccggcgc ggccccgatg ccccgggaac    102960
cacggagttg gtcgccgagc tgacggcgca cggtgccgag gtggccgtgc aggcatgtga    103020
cgtggccgat ggcgatgcgg tggcggcgtt ggtcgccggc gtgccggatg agcacccgct    103080
gagggcggtc gtccacacgg ccggtgtgtt ggacgacgga gtgatcggct cgctcaccga    103140
ggagcggctc gccaccgtcc tgcggcccaa ggcggatgcc gcctggcatc tgcacgaagc    103200
gacccgcggc ctggacctgg acgcgttcgt cgtcttctcc tccgtcgcag gggtcttcgg    103260
cggcgccggc caggccaact acgccgcggc caacgccttc ctggatgcgc tgatggccca    103320
gcgccgggca gcgggcctgc ccggactctc gttggcctgg gggccgtggg accagaccgg    103380
cggaatgacg ggcatgctgt cggacgccga ggccgaccgc ctcgcccgct ccggcatccc    103440
gccgctctcc gcggagcagg gcctcgccct cttcgacgcg gcactcgctc ttgccggaac    103500
cagcacgcca gacagggcag ccggcagcgc cgccgccagc acgtcgggaa ccggcgacac    103560
gatcgccatc ccggccgcgg ccctcgtcgc accggtccgg ctcgacctgg cggcgctggc    103620
cgcgcagggt gaggtccccg cgattctgcg cggactggtg cggacccgca cccgccgtac    103680
ggcggccggc ggttccgtca ccgtggccgg actcgtcaac cgcctgtccg ggctcaccgc    103740
cgacgagcgg cgccaggaac ttctcgaact ggtccgcact caggcagccc tcgtcctcgg    103800
```

```
gcacgccgat ccggcgtccg tggactccac cgcacagttc cgtgacctcg gcttcgactc   103860
gctgaccgcc gtcgagctgc gcaaccggct gagcacggcc accggcctgc gcctgaccgc   103920
aaccctggtc ttcgactacc cgaacaccga tgccctcgcg gagcacctgc gggacgagtt   103980
gttcggcgcg gtcgagagcg aggtgcgggt gccggtccag gcactgccgc cgaccgccga   104040
cgatcccatc gtggtggtgg gcatggcctg ccgtttcccc ggtggtgtga cctcgcccga   104100
ggacctgtgg cgcctggtcg acgccggcac cgacgccatc accaccttcc cgaccaaccg   104160
cggctgggac ctcgaatcgc tctacgaccc ggaccccgca cacctcggca cctcctacac   104220
ccgctccggt ggcttcctgc acgaggcggg ggagttcgac ccggcgttct tcggaatgag   104280
cccgcgtgag gcgctggcaa ccgactccca acagcgtctc ctgctggaat cctcctggga   104340
ggcgatcgag cgggccggca tcgacccgct gaccctgcgc ggcagcgcca ccggcgtctt   104400
cgccggcgtg atgtacagcg actacgggag catcctcggc ggcaaggagt tcgagggctt   104460
ccaaggccag ggaagtgcgg gcagcgtggc ctcgggccgc gtctcctaca ccctcggctt   104520
cgaaggtccc gccgtcaccg tggataccgc ctgctcctcc tccctggtcg ccctgcacct   104580
ggcagcccag gcccttcggg cgggtgagtg cacgctcgcg ctcgccggtg gtgtgacggt   104640
gatgtccacg ccaggcacgt tcgtggagtt ctcgcggcag cggggtctgg cgcctgatgg   104700
tcgttccaag gcgttcgccg aggccgcgga cggcgtcggc tggtccgagg gcgtcggcat   104760
cctcgtcctg gagcgccagt cggacgccgt gcgcaacggc cacgagatcc tcgccgtgat   104820
ccgcggctcg gcggtcaacc aggacggtgc gtccaacggc ctgaccgcgc ccaacggccc   104880
gtcccagcag cgcgtcatcc gtcaggcgtt ggccagtggc ggcctgtcca cggccgacgt   104940
ggacgccgtc gaggcgcacg gcacgggtac gacgctcggt gacccgatcg aggcccaggc   105000
gctcctggcc acctacggcc gcgaccgcga ccccgagaac ccgctgctgc tcggttcgat   105060
caagtccaac ctcggccaca cccaggcagc ggccggtgtc gccggtgtca tcaagatggt   105120
catggcgatg cggcacggcg tgctgccgca gaccctgcat gtcgacgcgc cgtcctcgca   105180
cgtcgattgg agcgtcggcg ccgtcgaact gctcaccgag cagaccgtct ggccggagac   105240
cggccgggtc cgtcgcgccg gtgtctcctc cttcggcatc agcggcacca acgcccacgt   105300
catcctggaa cagccggagg ccgtgcagcg cctggcaccg ggagcagcag agaccgtgga   105360
gccggtcgcc atcaagccgt cggcggaacc gtccctggtg ccgtgggcgc tgtccggcaa   105420
gtcacccgag gccctgcgcg cccaggccgc acgcctccgt gacttcctgg cggaacggcc   105480
cgaaccgcgc tcgatcgaca tcggccactc actggccgtc acacgctcgc agttcgacca   105540
ccgcgcgatc gtgctggtcg acgatgcgaa ggccccggcg gacagcctgg ccgccctcgc   105600
ggccctggcc tccggtgtgg ccgatcccgc cgtcgtctcc gacgcggtat cgaccggcgg   105660
ttcggcagtg ctgttcacag gtcagggtgc tcaacgtctg ggcatggggc gtgagttgta   105720
cggccgtttc ccggtcttcg ccgaggctct cgatgtcgtg gtcgaccacc tggacgccgc   105780
cttgcccgcc caggccggtt tgcgtgaggt gatgtggggc gacgatgtcg agttgctgaa   105840
cgagacgggt tggacccagc ccgcgctctt cgccgtcgag gtggcgctgt ccggctggt    105900
ggagcgttgg ggtgtccgtc cggacttcgt ggccggtcat tccatcggtg agatcgcggc   105960
ggcgcatgtc gccggggtgt tctcgctgga ggacgcctgc cgtctggtgg ccgcccgtgc   106020
gacgctgatg caggcgctgc cgaccggcgg cgcgatgatc gcggtccagg ccaccgagga   106080
cgaagtgacc ccgcacctga ccgacgaggt ggcgatcgcg gccgtcaacg gcccgacctc   106140
cgtggtgatc tcgggcgcag aagaggccac gcagaccgtg gcacaacact tcgccgacca   106200
ggggcgccgg acgaccgcgc tgcgggtctc gcatgcgttc cactcgccgc tgatggatcc   106260
gatgctggcg gagttccgtg cggtggcgga aggactgtcc tacgccaccc cgtccctccc   106320
cgtcgtctcg aatctgacgg gctggctggc cacggccgac gaactgtgct cggccgagta   106380
ctgggtgcgc cacgtccgcg aggcggtccg cttcgccgac ggcatcacca ccctcgaagc   106440
cgagggcgtg cggaccttcc tggaactcgg cccggacggc atcctgtccg cgctggctca   106500
gcagtccctc gccggcgaag ccgtcaccgt gcccgtcctg cgcaaggacc gcggtgagga   106560
gtccacggcc ctgacggccc gagcgcatct ccacacccgc ggactgatcg aagactggca   106620
ggacttcttc gctggtgtgg gtgcggggcg ggtggagttg ccgacgtatg cgttccagcg   106680
tgggtggttc tggccggttg gtcgggttgg tgttggtggt gatgtgggtg ctgtggggct   106740
tgggtctgcg gggcatccgt tgttgggtgc tgcggtggag ttggctgcgg gtgcggggt    106800
ggtgttgacg ggtcgtctgt cgttgtcgtc gcatggttgg ttggctgatc atgcggtgat   106860
ggggcgggtg tttgctcctg gtacggcgtt gctggagatg gtgatgcgtg ctggtgatga   106920
ggtggggtgt ggtcgtgttg aggaactgac gttggcggcg ccgttggtgt tgcctgagcg   106980
tggtggggtg cgggttcagg ttgctgtgga tgctcctgat gctgcgggtc gtcgtggtgt   107040
ggggtgtat tcgtgtcctg atggtgtggg tcaggcggtg tggtcgcagc atgctgtcgg   107100
tgtgttggcc tctggtgcgg ctgaccaggt cggtgggttc ggtgacggtg gtgtgtggcc   107160
gccgcagggt gcggtgtcgg tggatgctga gggctgctac gagctgtttg cggatgctgg   107220
gttcggttat ggcccggtgt tccaggggtt gcgtgcggtg tggcgtcgcg gcgaggaact   107280
cttcgcagag gtcgccctgt cggacgaggt tgctgagagc gctgatacgg cgaccggttt   107340
cgggttgcac ccggcgttgc tggatgcctc gctccatgcc tcgcttctct cctcccttga   107400
```

```
aggtcaatcc gccgatggtg ggcctgcgtt gccgttcgcg tgggagggtg ttttcctctt  107460
cgcctcgggt gcgacggctt tgcgcgtgcg gttggcgccg gcgggtgagc atgcggtgtc  107520
ggtgaccgcg gtggatccga ccggtgcgcc ggtgatttcc atcgacgcgc ttcgtacccg  107580
tcgcctcacc ctcgatgagg tcaacgcatc ccacacccag ctgagcgatg cgctcttcgg  107640
cgtccaatgg accacggtcc cgagcacccc ggccgccgac cacccgtcgg tagccatcat  107700
cggaaccgac cccttcggcc tcgcagacgg cctttcggac gccttgcccc tggtcgagga  107760
gcgcggtgac ctcgcggcgc tcgcagcgtc ggagcacccg gtaccggacc tcgtcctcgt  107820
cccggtagcg ggcacccggc gcacaggcgt acctgcggac gccgaaggac acaccgacgc  107880
cgggacatcc gacatgctcc gatccgtgcg tgaggccacc gcacaggtac tggagcagat  107940
ccagcagtgg ttggcggacg accggttcga ggcggcgcgg ctggtgttcg tgacgcgcgg  108000
ggcggtttcc gtgggtgagg gcggcatcgc cgacctggcg gcctcggccg tctggggtct  108060
ggtgcggtcg cgcagtcgg agaatccggg ctgcttcggt cttctcgacc tcgacctcga  108120
cctcgccctt gactccgacc ttgcccccga ggtcgacatc gagcgcgacc gtgaccgcga  108180
tccggtcggt gggaccgtgc agcccgcgct cgccgcggcc ctgcacgcga ccgccgacga  108240
gccgcagttg gcactgcgcg gcgggaccgt gcaggccgcc cgactgaccc gaatccccgc  108300
gccgcagacc gaccgtgccg agaccgaccc tgccgagacc gaccgtccgg agatcgacac  108360
ccggcggccc ggcacggtgc tcatcaccgg tggtaccggt ggcctcggtg ggttgctcgc  108420
ccggcacctc gtcgccgagc gggggtacg gagcctggtg ctcgccagcc ggagcggtct  108480
cgcggccgag ggagcggaga agctggtcgc cgacctcgaa gcgctcggtg ccgtggtggc  108540
cgtgcagacg tgtgatgtgg ccgatggcga tgcggtggcg gcgttggtcg ccggcgtgtc  108600
ggacgagtac ccgctgacgg cggtcgtcca cacggccggt gtgttggacg acggagtgat  108660
cggctcgctc accgaggagc ggctcgccac cgtcctgcgg cccaaggcgg atgccgcctg  108720
gcatctgcac gaggcgaccc gcgatctgga cctggacgcg ttcgtcgtct tctcctccct  108780
cgccggcgtc ctcggtggcg ccggtcaggc caactacgcg gcggcgaaca cgttcctgga  108840
cgccttgatg gcgcagcgtc gcgccgccgg gttgccgggt gtgtcgctgg cgtggggtcc  108900
gtgggaccgg gccggcggca tgacggggac cctgtcggac gccgaggccg accgcctcgc  108960
ccgctccggt gttccgccga tctcggcgga gcagggcctt gcgctgtacg acgcggcgac  109020
cgccggtgag cggccgctgg tggtgccggt gcggctggac ctcgccgcgc tccgcgggct  109080
cggtgatgtc ccggcgctgc tgcgcggact ggtccggacg cccgcgcggc ggaccgcggc  109140
ggccggtgcg gcgccgtcgg ccgatgtgct cacccggcag ttggccgggc tcggcggggc  109200
ggagcaggag gaggtcctgc tgaggctggt gcgcggtcag gccgcggtgg tgctcgggca  109260
cgccgacggc tcggcgatcg gtgcggggcg acagttccag gagttgggct tcgactcgct  109320
gaccgcggtg gagttccgca accgactcaa cgcggccacc ggactgcggc tgccggccac  109380
cctgctgttc gactacccga cgccggccga cgtcgtcggg cacctgcgcg ccggctcgg  109440
caccggggag gtgtcgggtg cgggctcggt gctggcggcg ctggacaacc ttgaggcggt  109500
gatcgccggg ctgtccctcg acgacgcggg ggagcaccag ttggtggccg gccggctgga  109560
ggtcctcagg gcgaagtggg cggacatgcg aagcgcggag ggagctgtgg acggcggtgc  109620
ggacgtcgac atcgaggagg cgtcggacga cgacatgttc gcgctgctgg acgacgagct  109680
ggggctgaac tgagccgctc cgcatgagca gttccgcacc agaagttcca cagtggcctg  109740
gtccacagtg acctggtcca cagtgagctg ttccacagcg acctgttccg cggcggttcc  109800
gtagtgagcc gttccgctta ggcgggaaac cggggccccg gtggtcggac gcttgattcc  109860
ggccaccggg agcgtcaaac cggcctcttc taaagaaagg gaaagaacta gggaaacacc  109920
cacagcccga tcatgtgcaa tgaattccgt ggctcggaaa accattcgcg ggccatggag  109980
ttacggtgtg atcacgtgtc cgtacggcag tgacggcgtg cacgggttgc cacggcaggt  110040
ccccggccga ggcgtctccc cgctgctgcc cggcggtgcg tgccgcttgg ttcgtcccta  110100
ggaggtcccc ccatgaccac gtccaccgag gagagcctgt gggcccggtg cttccatcca  110160
gcaccggccg cccccgtccg gctcttctgt ttcccacatg cgggcggctc ggcctccttc  110220
tacttcccgg tgtcggccca actgtcctcg gttgccgagg tgttcgccat ccagtacccg  110280
gggcgccagg accggcgcaa ggaagccggt gtcagtgacc tcgcgacctt ggccgaccag  110340
gtctacgacg cgctgcgccc cctgctgaag gagcggccga gcacgttctt cgggcacagc  110400
atgggcgcga cgctggcctt cgaggtggcc cggcgcttcg aggccgacga cggtgacctg  110460
gtccggctgt tcgcctccgg gcgccgggcc ccctcccgcg tgcgtgaaga ggccgtgcac  110520
cggcggtccg acgacggcat cgtcgaggag ctgaagctgc tcgccggcac caacaccgcg  110580
ctgctcggcg acgaggagat cctgcggatg atcctgcccg cgatccgcag cgactaccag  110640
gccatcgaga cctaccgctg cccgccccgac gtcaccgtcc gggcgccgct gaccgtcctc  110700
accggcgacc gcgacccgaa gacctccctg gacgaggccg aggcgtggcg cggccacacc  110760
accggggact tcgacctcaa ggtgcttccc ggtgggcact tcttcgtcag ctccgaggcc  110820
ccggcgatca tcgatctgct ccgggcgcac ctcgccggca acggctagcg ggcgcactgc  110880
ggcaggccgg cggtgccggt ctgccgcacg gcccccgcgc cgcctgagac ggcaccatgc  110940
cacccggcga cgcgcgttgc gtgcgcctgt ggagcaccgt cccgcgtgcg tacgcgggcg  111000
```

```
cgcggaggac gtgctcggcc ggccggatgt cctggaggag accattcgga aaacgtcaaa    111060
tgtctgacag ctgctgctca tttactccag cccgcactat atgatcttgt ggtggggcat    111120
gtacggagcg tgccactcgt tcgctcctgg tgcatccatc ccgaccgtcc ggatttgtgt    111180
gcaccggacg gccgtggtcc tggtacgcgt cggcgcgagg tcggctcctg ttctcacccc    111240
acgtggagga acgcgtgatg cggaagcagt cgggttcttc cggcttactg accacgctgg    111300
tgggacgcga cgacgaactg cgcaccctcg cccggcacgc cgcggccgcc cgcgacgggc    111360
gggccggcct ggtcctgctc cacgggcccg ccggcatggg caagaccagt ctgctgcggt    111420
ccttcacggc gagcgatgtc tgccgcggca tgacggtgct gtacggcacc tgcggcgaga    111480
ccgtcgccgg cgccgggtac ggcggtgtgc gcgaactcct cggcgggctc ggcctgagcg    111540
gcggcgacgc ccggcgctcg cccctcctgg agggcctggc ggcccgcgcg ctgcccgcac    111600
tcaccgcaga ccccgccggt cccgacgccg ccacgggtgc ctacccggtg ctgcacggcc    111660
tgtactggct cgccgcccgc ctcatggccc aacggccgct ggtcctcgtc ctcgacgacg    111720
tccactggtg cgacgaacgc tccctggcct ggatcgactt cctgctgcga cgcgccgagg    111780
acctgccgct gctggtcgtc ctggcctggc gcagcgaggc cgaaccggtc gcgcccgcgg    111840
tgctcgccga catcgccgcc cagcgccgcc ccaccgtgct cggcctgcac ccgctcggcc    111900
ccgacgacat aggcgaaatg gtgcgtcgcg tcttccggac cacggccgca ccgtcgttcg    111960
tcagccgggt cgccgccgtg tccggcggca atccgctggc cctcgcccgc ctcctcgacg    112020
aactccgcgc cgagggcgtc cggccggacg ccgccgggga gcgccgggcc gccgaggtcg    112080
gcagtcacgt cctcgcccgc tcggtgcgct gcctgttgga gcgccggccg ccctgggtgc    112140
gcggcgtggc ccgtgccatc gccgtactcg gcccggagtg caccgagttg ctggcggcgc    112200
tcgccggcgt cccggccgcg accgtcgacg aggccctgtt ggtgctgcgc agggccggca    112260
tcctggccgc cgaccgcgtg gacttcgtcc atgacgtcgt ccgctccgcg gtgctcgacg    112320
acgtcgcccc gcccaccctg gccgaactgc gcaccaacgc cgcgctgttg ctgagcgacg    112380
ccggccgccc ctccgaggag ctcgccggcc agctcatgct gctgccggtg ctcgaccagc    112440
cgtggatggc cgcggtgctg cgcgacgccg ccgcccaggc ggagagccgc ggcgccccgg    112500
aggccggtgt gcgctgcctc taccgggtgt tggaggtgga gccggacaac gttgccgtcc    112560
gcatccagat ggcccgcgcg ctcgccgaga tcaacccgcc cgaggcgatg cgcctcctca    112620
aggaagcgct ctccctcgcc ggggacgtcc gcacccgcgc ccaggtcgcc gtccagtacg    112680
gcttcacctg cctcgccgtg caggaatccc cgtccggggt gcggatgctg gaggacgcgc    112740
tcgccgagct gaccgccgaa ctgggccccg aaccagggcc cgtggaccgg gagttgcgga    112800
ccctcgtgga gtccgtgctg ctcatcgtcg gggccgacga gaaggtgacg atcggtgcgg    112860
tccgcgaccg ggcggcccgg ctcaccatgc cgcccggcga cacaccggcc cagcggcaga    112920
tgctggccat gaccaccgtg ctgaccgcga tggacggccg ggacgcccgg tcggccgtcg    112980
accaggcccg ccgcgccctg cgcgcacccg gcgtcgagct ggaaccctgg tcgctgctgt    113040
ccgcctcctt cgccctctcc ctggccgacg aggtcgccga cgcgcagtac gcactggacc    113100
tcatgctcca gtacggccag gacaacgcgg cggtgtggac gtacgtcctg cgcgctctcca    113160
cgcgcgccct gctccaccac ggggtgggcg ccttccccga ggccctggcc gacgcccaga    113220
ccgccgtcga gatcctcggc gaggagcgct gggcggacgg cgccgtgctg ccccgtgtcg    113280
cgctggccac cgccctcgtc gaccgcggcg agcccgagcg cgccgaacac gtcctcgacg    113340
gcatcacccg ccccgcctc gaacgcttcg tcatcgaata ccactggtac ctccaggccc    113400
gcgcctacgc ccgctgggtc cgcggggatt tccaaggagc cctggacctc cttctcgcct    113460
gcggtcggtc cctggaggag tcgcgcttca gcaacccggc gttcgtgccg tggtggcccg    113520
atggcgcggt gctcctggcg accctggacc gccacgacca ggcgcgcgaa ctcgccgcat    113580
acggaagcga gttggccgag cgttggggga cggcgcgcgg cctcggactg gccttcatgg    113640
cccagggcgt cgccgcaccc ggccgcgccg gcatcgatca cctcaccgag gcggtctcgc    113700
tgctcgcgga ctccccggcc cgggccatgg aggcccgggc cgaacttctt ctcggacacg    113760
cccacctgaa gcgcgacgac ctgcgggccg cccgggaaca cctgcgcgcc gccgccgacc    113820
tcgcccagcg ctgcggcgcc gtgaagctcg gcgtcgacgc cagaaaactg ctggtcaccg    113880
cgggtggtcg ggtacgcagg atgaccgcct ccccactcga catgctgacc gggatggaac    113940
gcacggtggc ggacctcgcg gtgaccggcg cgagcaaccg ggccatcgcg gaagccctct    114000
tcgtgactgt aaggaccatc gaaacccatc tcacgagcgt ctaccggaag ctcggggtcg    114060
gcgggcgtgc ggagctgtcc gccgtcctgg agaccaggac cgccacctcc ggtcggcagc    114120
cgccggcctg ggtctcccag gcacgcggac gcgcttgaga cgaacaggac gagaggtagc    114180
ggtgccccgg agcaaggcgc ggaaccaacc gaccacctgc acgccgcagt gcgcacccga    114240
cgcccacggc gaccccacca tgctcctcga atgcggccgg gaacagcggc tcatcggcga    114300
cctcctgcac cgcctcggcc aggggcggcc atcggtgctc agcctgaccg gccggcccgg    114360
gcacgcccag aacgccctgg tccgctgggg cgcgtgccgg gccaggcacg acgggctgcg    114420
cgtcctgcga gcccaggcga cgcccgcgga acggaactc cgctacggcg ccgttctcca    114480
actgctggcc gtcctcgacg gcccgcacgg cagcaccctg gacgccgcga tccgccacga    114540
cggtcccccg ccactgcccg tgcccggcat cgaggaggtg ctgcggcgca ccggcacggc    114600
```

```
acccaccctg gtcgtggtcg aagacgtcca gtggttggac ccggcctcgc tgacgtggtt    114660
gcagatcctg ctgcgccacc tcgggccgga caccccgctc gccgtcctgg ccagcagctg    114720
cggtgacacc acggccttcg acaccgaccc gaaggccccc gccgtcccgg ggccgccgga    114780
caccgtgccc gtcgcgcgct tcgtggtgcc cgcgctcacc gaccgcgggg tcgccgccac    114840
cgtccgcgcc gtctgcggca cccccggcga cgaggagttc atcgccgcgc tcacctccgc    114900
caccgccggc aaccccgcca tcctgcggga cgccctgcgc gccttcgtcg accacggcct    114960
ccccgccgac gccgaccacc tcccggagct gcacgccctc accgctggcg tcgtcggcga    115020
ccacaccgtg cgcgccctgg acggcctgcc cgccgaagtc aacgccgtcc tgcgggccct    115080
ggccgtctgc ggcgacctgc tcgacttcca ccgagtccgg gccctcgccg cgcgcactc     115140
gctgtccgag gaccggatcc gcaccctgct ggcgagcgtc ggcctgaccg tgtccgtcgg    115200
cgacaaggtg cacatccgct tcccccgcctc caaggcacgc gtcatcgagg acatgcccgc    115260
cgcggagcgc gccgatctgt acgtccgcgc ggccgaactc acccacagtt gcggcgtcaa    115320
cgacgaggac gtcgcccatc tgctgctgcg ctcgtcgccg ctcggcgcac cctgggtcgt    115380
gcccctgctc cgccgcggat tcgccgccgc gctgcgccgg gaggaccacc accgggcctg    115440
tgcctgcctc tcccgcgccc tgcaggaacc cctcgacccc cgggaacgca gcctgctgac    115500
gttggaactg gccgcggccg aagccgtcgc ccggccggag gcggggggatc gacgcctggg    115560
ggaactcgtc cgcagcaccg tcgcggacac cgaccccacg tcgtccggtg aggggtgtggg    115620
ggtccgcgcc atcgacctgg ggttcgcccg gggcaacagc gaatgggtcc gccgcaccgc    115680
gggcgaggcc ctgccgtacg ccgggccggc cgaccgggag gaactggtcg cgctgttctg    115740
gttggccgcc gtgcgggacg acgacgcgcc gatgatcccc gtggtgcccc ggttgcccga    115800
ccggccggtg ccgccggccc aggccggcgc ccgtgcctgg cagctggcca cggcgggga     115860
ggacgcggac aaggccagga agctcgcccg gatcgccctc accggcgggg tgaacgagag    115920
cctgatgatg. ccgaaactgg cggcctgcgc cgcgctgttc gccaccgacg acaacgacga    115980
ggcggtgcac ggcctggaca ccatgctcac cgccgcccgc agtgcccacc tgcgcagcat    116040
ggccgcccgc attttcaacc tacgggcccg gatacacctg tgcgcggccc ggctggaggc    116100
cgccgaacgc gatctggaca gcgccgagcg cgccctgccg ccgacgagtt ggcacccccg    116160
tgcgctgccc aacctgatcg ccacccgcat cctcgtcagc atggagacgg ccgcgcccgga    116220
ccgggcccgc cgactcgccg aggccccggt ccccgccggc ggcgaggagg gtgtgtggtg    116280
gcccgccctg ctgctcgccc gcgcccgtgt ggccgccgac gacggtgact gggaggaggc    116340
cctgcggctg tcgcggggagt gcgggcgctg gctttttcgc cggcactggg ccaacccggc    116400
catgctcagt tggcgcccgc tggccgccga ggcgtgtctg aagctcggcg acgtgacgga    116460
ggcgcgccgg ctgcgggacg aggagctgtt cttcgccgac cgctggggca ccgcgagcgc    116520
ccgcgggatc gcccgcctga cgacgcggcg actcttcgac gacgacggcg accgggccgt    116580
ccggcggatc cgcgaggccg ccgccctgct ccgcgactcg cccgcccgcc tggcctacct    116640
gtggagccgg ctgagccagg ccggtgccga gacggcccac ggcgacaccg ccgcggccgc    116700
acgctcctgg caggcggtcg cccggatgac cgccgcccac cccgccagcc gcctcgccac    116760
cgccgcccgc accctgaccg tcccgtccgt tccggtcgcc accgcgccgc ccaccgccgt    116820
cgtcccaccc ggatggcgcg acctgtccga ggcggagaag gacaccgtgc tgctcgccgc    116880
ccgcggccac ggcaaccgcc agatcgccga acaactcgcc gtcagcaggc gcaccgtgga    116940
gctccggctg agcaacgcct accgcaagct gaggatcggc ggacgcaagg agctgtacct    117000
gctcctggag gcgctggaag accggtcgc ggatgcttct tgagcgggag aacgaactgg    117060
cccggatccg ggccgccctg gacgccgcgg aagcgggcga ctcctcgctc ctcctgatca .  117120
acggtccct cggcagcggg cgttcggcgc tgctgcgccg gataccggag ctggccggcg     117180
acggcacccg cgtcctgcgg gccagcgccg cctggcggga acgcgacttc cccttcggga    117240
tcgcccgcca actcttcgac cacctgctgt ccggggcggg cggcgcaggg ccggccgaac    117300
gcaccgccgg ggcagagcac ttcagccgac tgatggacac cggcgaccgc cctaccggga    117360
ccggccccgc cctggaggtc tcccaggcag tgctccaggg cgcccaggcg ctgctcgccg    117420
acgcgtccgc ggagcggcgc ctgctgatcc .tggtcgacga cctccagtgg gccgacggcc    117480
cgtccctgcg ctggctggcc cacctcaccc ggcggctgca cggcctgcgg gcgctgctgg    117540
tgtgcacgct ggccgacggc gaccaccggg gcaggtaccc cctggtccgg gaggtcgccg    117600
gcgccgcgca caccgtcctg cgcctggcgc cgctgtcccg ggacgccacc cgcgtcctgc    117660
tcgccgggcc ccagggccgg ccgccgcagg acgcactggt gcgcgccgtg tacgaggcgt    117720.
ccaggggcaa cccgctgttc ctgaccgcct tccggagcgc tctgcgcgcc accggaaggc    117780
cgcccggcgg cgaccacttc ggcgccgtcc gggagctgag cccgacggtg ctgcgcgatc    117840
ggctcgcggg ccatctgcgg atccagccgc agccggtgcg cgaggtcgcg gtggcggtgg    117900
ccgcgctggg cgaccacagc gatccggtgc tgctcgccca gctcgccggg gtcgatgaaa    117960
tcggtttcgc cggtgcccgc cgcgcgctgg tggacgccgg cctgttggcc cggggacggg    118020
acgtccgctt cgtccacggc gtcgtccgcg atgcggtgga ctccctgctc accctcgacg    118080
agcgggaacg ctcgcacgac gacgccgccg atctgctgta ccgctgcggg cggccggccg    118140
agcaggtcgc cggccatctg ctggccgtgg tccacccggg ccggccctgg tcggaggcgg    118200
```

```
tcctgcgctc cgcggcccac aacgcgctgc gcgccggccg gcccgccgac gcggcccggt   118260
acctgcgccg cgccctgctg caccaccgca cccaggacgg ctgccgcgcc cgcatcctgg   118320
tcgatctggc caccgccgag cgcgccctcg accccgatgc ctgtgtacgc cacgtcagcc   118380
aggcggtcgc gctgctggac acctcgcggg atcgggccgc cgccgtgttg cgcatcccgc   118440
cgtccctgct cgccgccccc agcccgtccg ccgtcgagtt ggtgcggcag gccgccgccg   118500
ggctcgacga accggggcag cgggacgagg agggagccga cgaactcgcc ctgcgcctgg   118560
aggcgtggct gcggcactcc ggccacgaga accccgtcga gctggcgtcc tcggtggcgc   118620
ggctgcggcg catgggggca cggccgccgg tggacagcgt cgccgaacgc gagttggtcg   118680
ccgtgctgtt gagtgcgggc gcgctcagcg gccggctcag cgccgcggag atcgccgaca   118740
ccggcaaccg catcctcgaa cgtgagccgg ccaccgccgc ccatgcccac accccgctgc   118800
cgctggtgat gctctcgctg ttcgtcgccg agtccgtgca gggcgtggcc tcctggctgg   118860
ccagcgaaca gcacacccgg cgccggtacg cgaccggcgc ggacgatgtg ctgctgaccg   118920
ccgagcgggc cttcgtcctg gtgacgcagg gccgcccggc cgccgcgcgg gagcacgtcg   118980
aacgcgccct ggtcatggac gccggcgact ggtcggaacc cgccgtcatg atgttcgccg   119040
cggtcgcctt cgaactgcgc gacccggcct tgagcgaacg catcctggaa cggatccgcg   119100
accgccggcc ggccggactg gcgctcaccg ccaccggtca gatgctccag gccgccgtcg   119160
acgtgcactt cggccggggg cgggacgccc tggacacgct gctggcctgt ggccgacgcc   119220
tggagaccgt gggatggcgc aattccgcac tgctgccctg gcgtccgtat gcaatcgggc   119280
tgcaccagcg gctcggcgag accgatgccg cgctgcaact cgccgaggac gagctgaggt   119340
gggcccggga gtggggtgcg acgacgaacc tgggccgggc cctgcgcctg aagggctggc   119400
tgttgcagga cgagggactg gatctgctgc gtgagagcgt cgagatcctc cgcgcttcgt   119460
cctacgcgac ggaactggcg cgcaccctcg tcgtcctcgg gcgtcggctg ccgggtgggc   119520
cggaggccga ggcggtgttg cgggaggccg cggggatcgc cgccgcctgt ggtgttccct   119580
ggctcgccga acgggccgaa ctcggtctgg gcagtgccat cgtgccgccg gtcgccaccc   119640
tgaccccccag cgagcggcga gtggcgtcgc tggtgagccg gggcctgacc aaccaggcca   119700
tcgcgaccga actcggtgtg agctcccggg cggtggagaa gcacctcacc agcgcctatc   119760
gcaagctcgg cgtctccggc cgccgcgagt tggtgaatgc gctcccgggt cgttgacgcc   119820
cgcggccgtc ccgttccgga cgatcacaat cctctgtgcc tgtcagccct gtgggtgcgg   119880
cgtcgttgtc gtacgcccgt ccggtcatga cagcaatcct caacaaaagc tcaaacggaa   119940
gcggggggcag gtgcatccac cccatcacca atgtctcgta tgttactttt ggcccaccgt   120000
agttgggact ctgcggaccg tacctcgtca agtacggacc acggatttag gggtgcgtag   120060
atgtgtcgcg ttgacgtgcc gagttgtggc taactacgct ccgtttcacg gaaatcgaaa   120120
ctcttcggcg tgaatcgcgg cgaacttcga tcgaattctg gattccgttt ctcagaaggc   120180
taaaagacga cgggggttat ccgaaaggat ggcggtctcg tgactatcac tcacctcact   120240
gactcgaatc accgtacgac cggcggggtg atatccgctc agaccgcacc ggccggggag   120300
tccgtcggcc ccgggctgat ggcgtccctc gaccgtgacc tcaccatcaa gcacgccaac   120360
caggagttcc gccgccgctt cgacgattcc gccggtgacg tctgcggtcg cagcttccgg   120420
gacctgatgc acccgagcgt gcagcagccg ctgatgcgcc agttctcccg gctgatcgag   120480
ggcaagcggc accgcttcgc ctcgcacgtg gtcgccgtgg gcgcccagga cgcggccttc   120540
gcgggcaccc tcaccgcctc cgcggtcacc ggcaagaccc cggacatcgc cgggatcctg   120600
gtgctcatgg actcgtccgg cgccgccgac gccgccgatg ccggcgtcgt gaccagccag   120660
aagaagttcc tgaccgagat cgatgcgcgc atcctcgaag gcatcgcggc cgggctgtcg   120720
accattccgc tcgcctcgcg gctctacctc agccgccagg gcgtggagta ccacgtcacc   120780
gggctgctgc ggaagttgcg ggtgcccaac cgcgccgcgc tcgtctcgcg cgcctactcc   120840
atgggcatcc tgaacgtcgg cacctggccc ccgaaggtcg tggacgactt catcaagtga   120900
cgcccggccg gcctgtcccg caccggccgg ccaccgtcgg accgctgtcc cgcacgcccg   120960
tgtcgccgac gcccgcctct cacggggcga tgcccggcgg cacgggcgtt tcgtgccctg   121020
gtcgatgttt atccattctc acgttacgaa cccgcaatgc tcatgcgcgc gattccggcc   121080
acttttcgtt ccgtcgtaac caaactgccc gtcatatcgt ttccactgac cgtttctgtg   121140
gtccgcggtg tgccacctgg gccccactcc gacgcgcccg cgacgggtcg cgagcggcct   121200
ccgaatactc acggtaaccc ccgttcctcc tggtggagtt ccccgcgcgt cacattcctc   121260
cgtgcctcgt gcggcgtggc tcgcccgctt tccccgccgc cgcgacagcc attgttacaa   121320
cggcacgaca gagcggggggt ggtcgacccg tccgcggtgc ggcgcagtca agcggaacct   121380
ctgctcgatg ttcgtgccac cccgaactct gtgttcttgg tgaccgtttt caccaatgcc   121440
ccgcccgttc ggtggcgaat ccgccgcccg ttcctggcca caacggatcg cttcggccgc   121500
ccaccacggc cggcgcagcc gaggcttgtt ggaatcaccc ccacccagtc ggggcgccgt   121560
tgtggcccgg acacgcagga ggaacaccgt ggatgctgag ggccgccgca gggacatgct   121620
ggagctgatc cggcgcagcg gatccgccga tgtcgtgcgt ctcgccgagg agttcgccgt   121680
cagcaaggaa acggtccgcc gggacctgaa cgtcctggag ggccatgggc tgatacggcg   121740
gcggcacggt ggcgcctatc ccatggtgcg gccgggctcc gaggccgtct tcgtctcccg   121800
```

```
gaccgcacag ccgatcccgg aggagtcccg gatcgccacc gcggccgccg aactcctcag      121860
cgaggccgag acggtcttca tcgacgaggg cttcacccccg caactcatcg ccgacgccct     121920
gccgcgcgac cggccgctga ccatcgtcac cgcctcgctc ccggtggtca gcgccttcgc      121980
gacgagccca caggccaacg tgctgctcct gggcggccgg gtccgccggg gcacgacggc      122040
caccgtcgac cactgggccg tccacatgct gtccggcttc gtcatcgacc tggccttcct      122100
cggcgcggag gggatctcgc gcaggtacgg cctgaccacc cccgacccgg cggtcgccga      122160
ggtcaaggcc caggccatcc gcgtcgcgcg ccgcccggtc ctcgccgggg tgcacaccaa      122220
gttcggcacg gcgagcttct gccggttcgg agaggtgggc gacctggaga cgatcgtcac      122280
cggcgccggc ctgcccgtcg ccgaggccca ccgctaccac ctcatgggcc ccaaggtttt      122340
acgggtgtga cgccgccggg gcgtcccgcc gcccgcccac ggcctggcgc gcgccggtgc      122400
gggtcagccg cgtcgcgcgt ccgtcggttc ccgccagagt gcgagcggtg tcacccccac      122460
ccggtagccc tcctcgtcga ggagatggcc gccgtgcttc atgtgccaga gggcgaccgc      122520
gcgttcggtg acggtcaggt ccgcggcccg ttcggcgagt cggacctcga agcgctgggc      122580
gtcgtcgagg gagcgcaccc aggccaccag gtgcaggttg tgccggctga tgacgctcgc      122640
gcacaggcgg acctcgcgca tcccggtgac ccggcgggtc acctcgcgca gcctcgcggc      122700
cgggacctgc ccccagaagc tcaccgtgac cggccactcg gacagcgggc gggcgacctc      122760
gcaccgggcg tgcaacatgt ccgcggcgaa gagccgttgg acgcgacggc gcacggtgtc      122820
ggggccggcg ccgcactgct cggccagcgc gcggtaggtg gcccggccgt ccacggacag      122880
cgcggtgacc agctgttggt cgagctcgtc gaggacgaag gcggggggtgt ccgtgcggtg     122940
tctggacgcg tcggcggcga gccgggcgac ctggtgccgg ccgagggccc gcaggcgcca      123000
ccggctgccc tcggtgtgca ccgggccggc caggtgcgtc cgggcggccc ggacgccgtc      123060
caacgcggcg aggtcgtggg tcacccaacg ggagagcatg gccggatcgc gggccatgac      123120
gttgagttgg aggtcacggt ccccggtgac gtgggagagc gccaccacgt gcggcgcggc      123180
ggccaacgcc cgcgcgacgt cgagcagtcg gccgggagcg cagtcgatct cgacgaacgc      123240
caggcacccc tggcccgacg cggccagcac cggggccgga tggcagctga tccaggcggc      123300
gccggtctcg accagccggt tccagcgcct ggccaccgtc acggcgtcca gtccgaggac      123360
ggagccgatc cgggtccagc tggcccgcgg cgtgatctgc agcgcgtgca ccagggcctg      123420
atccacatgg tccagggacc gggggggtctg cccggaatcc tgcgccacgg gccacctcct     123480
tgcgtgtttc cggcggattc gggccgccgg tcggctcaac cttcagcctg gactcgggta      123540
cggccggacc gtaccaggca accccggag caacaggagt                             123580
//
```